Central Lancashire Online Knowledge (CLoK)

Title	Correlated orientations of the axes of large quasar groups on Gpc scales
Type	Article
URL	https://clok.uclan.ac.uk/40651/
DOI	https://doi.org/10.1093/mnras/stac269
Date	2022
Citation	Friday, Tracey, Clowes, Roger G orcid iconORCID: 0000-0001-8370-465X and Williger, Gerard M (2022) Correlated orientations of the axes of large quasar groups on Gpc scales. Monthly Notices of the Royal Astronomical Society, 511 (3). pp. 4159-4178. ISSN 0035-8711
Creators	Friday, Tracey, Clowes, Roger G and Williger, Gerard M

It is advisable to refer to the publisher's version if you intend to cite from the work.
https://doi.org/10.1093/mnras/stac269

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

Correlated orientations of the axes of large quasar groups on Gpc scales

Tracey Friday,[1]* Roger G. Clowes [1]* and Gerard M. Williger[2]

[1]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK*
[2]*Department of Physics and Astronomy, University of Louisville, Louisville KY 40292, USA*

ABSTRACT

Correlated orientations of quasar optical and radio polarization, and of radio jets, have been reported on Gpc scales, possibly arising from intrinsic alignment of spin axes. Optical quasar polarization appears to be preferentially either aligned or orthogonal to the host large-scale structure, specifically large quasar groups (LQGs). Using a sample of 71 LQGs at redshifts $1.0 \leq z \leq 1.8$, we investigate whether LQGs themselves exhibit correlated orientation. We find that LQG position angles (PAs) are unlikely to be drawn from a uniform distribution (*p*-values $0.008 \lesssim p \lesssim 0.07$). The LQG PA distribution is bimodal, with median modes at $\bar{\theta} \sim 45 \pm 2°$, $136 \pm 2°$, remarkably close to the mean angles of quasar radio polarization reported in two regions coincident with our LQG sample. We quantify the degree of alignment in the PA data, and find that LQGs are aligned and orthogonal across very large scales. The maximum significance is $\simeq 0.8$ per cent (2.4σ) at typical angular (proper) separations of $\sim 30°$ (1.6 Gpc). If the LQG orientation correlation is real, it represents large-scale structure alignment over scales larger than those predicted by cosmological simulations and at least an order of magnitude larger than any so far observed, with the exception of quasar-polarization/radio-jet alignment. We conclude that LQG alignment helps explain quasar-polarization/radio-jet alignment, but raises challenging questions about the origin of the LQG correlation and the assumptions of the concordance cosmological model.

Key words: methods: statistical – surveys – quasars: general – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

The spins of galaxies tend to align with the cosmic web of filaments, sheets, and voids. For example, Zhang et al. (2013) find the major axes of Sloan Digital Sky Survey (SDSS; York et al. 2000) DR7 galaxies are preferentially aligned with the direction of filaments and within the plane of sheets, and Tempel & Tamm (2015) find that orientation of SDSS DR10 galaxy pairs is aligned with their host filaments. Recently, Welker et al. (2020) detected a mass-dependent transition of galaxy spin alignments with filaments, from parallel at low mass to orthogonal at high mass. They found that this shift occurred at $10^{10.4-10.9}$ M$_\odot$, consistent with Horizon-AGN predictions (Dubois et al. 2014; Codis et al. 2018).

Cosmological simulations such as Horizon-AGN (Dubois et al. 2014) and SIMBA (Davé et al. 2019) predict the spin of dark matter haloes (and galaxies) are preferentially aligned with filaments and sheets at low masses (mainly spirals) and orthogonal at high masses (mainly ellipticals) (e.g. Dubois et al. 2014; Codis et al. 2018; Kraljic, Davé & Pichon 2020). Using the *Planck* Millennium simulation (Baugh et al. 2019), Ganeshaiah Veena et al. (2018) demonstrate this is a result of accretion history, with low-mass haloes tending to accrete mass from orthogonal to their host filament and thus orientating their spins along the filaments. In contrast, they find high-mass haloes tend to accrete along their host filament and have spins orthogonal to them.

The spins of quasars are also thought to align with the cosmic web. Large-scale alignment of quasar polarization was first reported by Hutsemékers (1998), who found the polarization of optical light from quasars was coherently oriented on Gpc scales at redshifts of $1 \lesssim z \lesssim 2$. This was then confirmed at higher significance levels by further polarization observations at optical wavelengths (Hutsemékers & Lamy 2001; Cabanac et al. 2005; Hutsemékers et al. 2005), the introduction of coordinate-invariant statistics by Jain, Narain & Sarala (2004), analysis using a new and completely independent statistical method proposed by Pelgrims & Cudell (2014), and polarization measurements at radio wavelengths (Tiwari & Jain 2013; Pelgrims & Hutsemékers 2015). Quasar-polarization alignment, although widely and independently reported, remains somewhat controversial (e.g. Joshi et al. 2007; Tiwari & Jain 2019). At least some of the controversy, however, appears to arise from different authors considering different scales and using different approaches to test for alignments (e.g. Pelgrims & Hutsemékers 2015).

Potential line-of-sight mechanisms for the large-scale alignment of quasar polarization must be considered. From the first detection, interstellar polarization was a concern but deemed unlikely (Hutsemékers 1998). More recently, Pelgrims (2019) finds the alignments are robust against Galactic dust contamination. Another potential line-of-sight mechanism widely discussed is exotic particles, such as axion–photon mixing in external magnetic fields (e.g. Cabanac et al. 2005; Das et al. 2005; Hutsemékers et al. 2005; Payez, Cudell & Hutsemékers 2008; Agarwal, Kamal & Jain 2011; Hutsemékers et al. 2011), although this is disfavoured using constraints from circular polarization measure-

ments (Hutsemékers et al. 2010; Payez, Cudell & Hutsemékers 2011).

If polarization is not induced along the line of sight, we must consider intrinsic alignment of the quasar spin axes (e.g. Hutsemékers 1998; Cabanac et al. 2005; Pelgrims 2016). Hutsemékers et al. (2014) report that optical quasar polarization is preferentially either aligned or orthogonal to the host large-scale structure. They propose that this bimodality is due to the orientation of the accretion disc with respect to the line of sight, and conclude that quasar spin axes are likely parallel to their host large-scale structures. Pelgrims & Hutsemékers (2016) report a similar result using radio wavelengths and large quasar groups (LQGs). They also conclude that the quasar spin axes are preferentially parallel to the LQG major axis for LQGs with at least 20 members, although they suggest this becomes orthogonal with fewer members ($10 < m < 20$).

Several studies report that radio jets are aligned over large scales (Taylor & Jagannathan 2016; Contigiani et al. 2017; Mandarakas et al. 2021), supporting the intrinsic alignment explanation independently of polarization measurements. [The work by Mandarakas et al. (2021) appears to supersede earlier work by the same group, Blinov et al. (2020), in which no alignment was found.] The potential correspondence of the QJARs (quasar jet alignment regions) from Mandarakas et al. (2021) with other large-scale structures such as the regions of correlated polarizations is a notable feature. In general, the corroboration of very large structures by independent tracers can provide compelling support.

In this paper we investigate for the first time whether LQGs exhibit coherent orientation, and whether this can explain the reported alignments of quasar polarization from Hutsemékers (1998) to Pelgrims (2019). This examines scales larger than those so far analysed, and potentially offers corroborating evidence for, and enhancement of, the intrinsic alignment interpretation of the results from many quasar polarization studies (e.g. Hutsemékers & Lamy 2001; Jain et al. 2004; Cabanac et al. 2005; Tiwari & Jain 2013; Pelgrims & Cudell 2014; Pelgrims & Hutsemékers 2015). If true, it would represent large-scale structure alignments over $\gtrsim$ Gpc scales, larger than those predicted by cosmological simulations and larger than any so far observed.

The concordance model is adopted for cosmological calculations, with $\Omega_{T0} = 1$, $\Omega_{M0} = 0.27$, $\Omega_{\Lambda 0} = 0.73$, and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. All sizes given are proper sizes at the present epoch.

2 DATA AND METHODS TO DETECT LQGS AND MEASURE THEIR ORIENTATION

2.1 Detecting large quasar groups

Our LQG sample is taken from the work of Clowes et al. (2012, 2013). The LQGs were detected using quasars from the SDSS (York et al. 2000), specifically Quasar Redshift Survey Data Release 7 (DR7QSO; Schneider et al. 2010). The DR7QSO catalogue of 105 783 quasars covers a region of ∼9380 deg², with its main contiguous area of ∼7600 deg² in the north Galactic cap (NGC).

Clowes et al. (2012, 2013) restrict their quasar sample to low-redshift ($z \leq 2$) quasars with apparent magnitude $i \leq 19.1$ in order to achieve an approximately spatially uniform sample (Vanden Berk et al. 2005; Richards et al. 2006). They further restrict their sample to a redshift range of $1.0 \leq z \leq 1.8$, within which the proper number density of quasars as a function of redshift is sufficiently flat for clustering analysis. They then detect LQGs using a 3D single-linkage hierarchical clustering algorithm, also known as friends-of-friends (FoF, Appendix A).

The resultant LQG sample[1] contains 398 LQGs. In order to confidently determine the geometric properties of the LQGs (e.g. orientation and morphology) we restrict their original sample to those with membership $m \geq 20$, giving a sample of 89 LQGs.

We select the most convincing of these using the significance estimates of Clowes et al. (2012, 2013). While the absolute values of these may be contentious (Nadathur 2013; Pilipenko & Malinovsky 2013), they provide a legitimate relative order for ranking based on confidence. As a compromise between sample size and confidence, we restrict our sample to LQGs with 'significance'[2] $\geq 2.8\sigma$, yielding 72 LQGs.

We finally exclude one LQG in the south Galactic cap, giving our final sample of 71 LQGs, of varied and generally irregular morphologies, as shown in Appendix B.

2.2 Determining large quasar group orientation

The position angle (PA) of a large quasar group can be calculated in either two or three dimensions. For our sample of 71 LQGs we find that the two approaches are generally consistent. We use the 2D approach, which involves tangent plane projection of the LQG quasars, followed by orthogonal distance regression (ODR) of the projected points. However, data from the 3D approach, which involves principal component analysis of the LQG quasars' proper coordinates, are used for some preliminary analysis of the morphology of LQGs. See Appendix C for details of both approaches.

Due to the filamentary nature of LQGs, orthogonal distance regression gives a better linear fit for some LQGs than others. Therefore, we have higher confidence in some PAs than others. We weight the PA of each LQG according to its ODR goodness of fit by inverse residual variance per unit length as

$$w = \ell/\sigma^2, \tag{1}$$

where ℓ is the length of the ODR line fitted to the LQG, and σ^2 is the residual variance of the m quasars in the LQG, calculated as

$$\sigma^2 = \frac{1}{m-1} \sum_{q=1}^{m} e_q^2, \tag{2}$$

where e_q is the orthogonal residual of the qth quasar from the ODR line. Note that this definition of weight (equation 1) is dimensionless only after normalization. Where possible we apply our statistical methods (Section 3) to both unweighted and weighted PAs.

We measure large quasar group orientation as the position angle from celestial north. It is important to recognize that the PA data are axial [0°, 180°), more specifically 2-axial; 0° and 180° are equivalent. In addition to analysing raw 2-axial PAs, some of our statistical methods (Section 3) require these to be transformed to vector (circular) data [0°, 360°). Following Hutsemékers et al. (2014) and Pelgrims (2016), we also test for alignment and, simultaneously, for orthogonality, using 4-axial data [0°, 90°). See Appendix D for details of these transformations.

We estimate PA measurement uncertainties using bootstrap resampling. For each of our sample of 71 LQGs, we create $n = 10\,000$ bootstraps and calculate their PAs. We find that the circular mean of the bootstraps generally agrees well with the observed PA, with a mean (median) half-width confidence interval (HWCI) of ∼10°

---

[1] Clowes (2016), private communication.

[2] We attribute no significance to the value of 2.8; it is used as a relative threshold only.

($\sim 8°$). See Appendix E for details of the bootstrap method, and calculation of HWCIs and their circular means.

2.3 Coordinate invariance: parallel transport

Position angles are dependent on the coordinate system in which they are measured, and in particular the position of the pole used to define $\theta = 0°$. To overcome this coordinate dependence we follow studies of galaxy spin alignment (e.g. Pen, Lee & Seljak 2000), CMB polarization (e.g. Challinor & Chon 2002), and quasar spin alignment (e.g. Jain et al. 2004) and use parallel transport.

For two objects at locations P_1 and P_2 on the celestial sphere, Jain et al. (2004) proposed parallel transporting the vector at the location of one object to the location of the other before comparing them. The path they use is the geodesic (great circle) between the objects. Parallel transport preserves the angle between the PA vector and the vector tangent to this geodesic. The correction to apply between P_1 and P_2 is the difference between the angles the geodesic makes with one of the basis vectors at each location. That is, if the tangent plane to the sphere has local basis vectors $(\hat{\theta}_1, \hat{\phi}_1)$ at location P_1, and the tangent unit vector to the geodesic at this point is given by $\hat{t}_1$, then the angle ξ_1 between $\hat{t}_1$ and $\hat{\phi}_1$ is given by (Pelgrims 2016)

$$\xi_1 = \tan^{-1}(-\hat{t}_1 \cdot \hat{\theta}_1, \ \hat{t}_1 \cdot \hat{\phi}_1), \tag{3}$$

with angle ξ_2 at location P_2 being similarly obtained.

The parallel transport correction between locations P_1 and P_2, i.e. the angle by which a vector rotates during parallel transport from P_1 to P_2, is given by the difference between angles ξ_1 and ξ_2 (Jain et al. 2004). So, to parallel transport the position angle θ_k of object k to the location of object i we compute

$$\begin{aligned} \theta_k^{(i)} &= \theta_k + \Delta_{k \to i}, \\ &= \theta_k + \xi_k - \xi_i, \end{aligned} \tag{4}$$

where θ refers to position angle, not spherical coordinates. Applying these corrections results in coordinate-invariant statistics (Jain et al. 2004; Hutsemékers et al. 2005). The result of parallel transporting a vector from P_1 to P_2 depends on the path taken between them. If a different path was chosen the parallel transport correction ($\Delta_{k \to i}$) would differ.

2.4 Mock LQG catalogues

To assess compatibility of the observed LQG PA distribution with that expected in the ΛCDM cosmological model we use mock LQG catalogues constructed by Marinello et al. (2016). They take a snapshot of the Horizon Run 2 (HR2) simulation (Kim et al. 2011) at redshift $z = 1.4$, and divide the volume into 11 sub-volumes. They then create quasar samples by applying a semi-empirical halo occupation distribution (HOD) model 10 times to each of the 11 sub-volumes. Finally, they use the LQG finder of Clowes et al. (2012, 2013) (Appendix A) to construct 110 mock LQG catalogues.

We restrict each mock catalogue to LQGs with membership $m \geq 20$ and significance $\geq 2.8\sigma$. The mean number of LQGs in our mocks is $\bar{n} = 30 \pm 0.4$; numbers in individual mocks vary $20 \leq n \leq 42$. We stack these to increase the statistical power. The 10 quasar mock catalogues created from each sub-volume are not truly independent (Marinello 2015); each HOD model realization samples the same set of dark matter haloes. Therefore, for each realization we stack the 11 sub-volumes, which are independent. The mean number of LQGs in each stack is $\bar{n} = 330 \pm 3$. The total number of LQGs in all 110 mocks is 3296.

We calculate position angles for mock LQGs as for our observed sample, including applying parallel transport corrections.

3 METHODS FOR THE STATISTICAL ANALYSIS OF LQG POSITION ANGLES

Statistical analysis of large quasar group position angle data requires appropriate methods. LQGs are widely and non-uniformly distributed, both on the celestial sphere (separation $\lesssim 120°$) and in redshift ($1 \leq z \leq 1.8$), and their PAs are axial data. Furthermore, the PA distribution may be bimodal, with PAs both aligned and orthogonal (Hutsemékers et al. 2014; Pelgrims 2016). Many statistical methods lack discriminatory power in multimodal cases.

We use methods for statistical analysis of the uniformity, bimodality, and correlation of LQG PAs, specifically to determine:

(i) Are they likely to be drawn from a uniform distribution?
(ii) Is their distribution bimodal, and where are the peaks?
(iii) Are they more correlated than random simulations?

3.1 Uniformity tests

For coordinate invariance, we perform uniformity tests on PAs parallel transported to the centre ($\alpha = 193.6°$, $\delta = 24.7°$, J2000) of the A1 region (Hutsemékers 1998) of large-scale alignment of the polarization of quasars, which is roughly at the centroid of the LQG distribution. (Alternative centres for parallel transport are discussed in Section 5.2.) We test the PA distribution for departure from uniformity using Kuiper's test, the Hermans–Rasson (HR) test, and the χ^2 test.

3.1.1 Kuiper's test

Kuiper's test (Kuiper 1960) is a rotationally invariant version of the better-known Kolmogorov–Smirnov (KS) test. It quantifies the maximum positive *and* negative differences between an empirical cumulative distribution function (EDF; our PAs) and a theoretical cumulative distribution function (CDF; in this case uniform).

To incorporate weighting we compute a weighted EDF, where for any measurement x, $F_{EDF}^w(x)$ is equal to the sum of the normalized weights of all measurements less than or equal to x. Following Monahan (2011), that is

$$F_{EDF}^w(x) = \sum_{i=1}^{n_x} w_i \Big/ \sum_{i=1}^{n} w_i, \tag{5}$$

where n is the total sample size, n_x is the number of measurements up to and including x, and w_i are their goodness-of-fit weights (equation 1). Kuiper's test is then computed normally, using $F_{EDF}^w(x)$ in place of $F_{EDF}(x)$.

The rotational invariance of Kuiper's test makes it independent of the 'origin' PAs are measured against (in this case celestial north). This makes Kuiper's test appropriate for circular and axial data that 'wrap' between one end of the distribution and the other, and also gives it equal sensitivity at all values of x.

The p-values are evaluated by simulation. We generate 10 000 samples of n random PAs, drawn from a uniform distribution, apply the same weighting, and calculate the fraction of samples with Kuiper's test statistic at least as extreme as the observations. Note that using alternative weights (e.g. w_i^2, $w = 1/\gamma$, $w = 1/\gamma^2$ where γ is the half-width confidence interval) does not significantly affect the p-value.

We apply Kuiper's test to both unweighted and weighted PA distributions of both 2-axial and 4-axial PAs.

3.1.2 Hermans–Rasson test

Landler, Ruxton & Malkemper (2018) test the performance of the Rayleigh and Kuiper's tests (amongst others) with a variety of multimodal distributions. They show that these tests lack statistical power in most multimodal cases, and find the Hermans–Rasson (HR) test for uniformity on the circle (Hermans & Rasson 1985) significantly out-competes the alternatives. The HR method is a family of tests, based on decomposing a circular distribution using Fourier series (Landler, Ruxton & Malkemper 2019). Variants of the HR test are controlled by the parameter β, with $\beta = 2.895$ being recommended by both Hermans & Rasson (1985) and Landler et al. (2018)[3] as offering power in both unimodal and multimodal cases. In this case, the HR statistic T of n measurements $\theta_1,..., \theta_n$ is defined (Landler et al. 2018) as

$$T = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi - |\pi - |\theta_i - \theta_j|| + 2.895|\sin(\theta_i - \theta_j)|, \quad (6)$$

where θ_i is the PA of the ith LQG and θ_j is the PA of the jth LQG, from a sample of n LQGs.

The p-values are evaluated by simulation. We generate 10 000 samples of n random PAs, drawn from a uniform distribution, and calculate the fraction of samples with the HR statistic T at least as extreme as the observations. Note smaller T statistics are more significant (opposite to KS and Kuiper's tests).

Our implementation of the HR test does not currently incorporate weighting. The HR test requires circular data so we apply it to 2-axial and 4-axial PAs after transformations $\Theta_{2ax} = 2\theta$ and $\Theta_{4ax} = 4\theta_{4ax}$, respectively (see Appendix D).

3.1.3 χ^2 test

The χ^2 test will have lower discriminatory power than tests applied to continuous data, such as Kuiper's and HR tests. Unlike those tests, it does *not* account for the 'wrap-around' nature of circular/axial data. It is included predominantly due to its ease of computation and interpretation.

For a histogram comprising m bins, the χ^2 statistic is

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}, \quad (7)$$

where O_i is the observed frequency and E_i is the expected frequency (in this case uniform) per bin i. To incorporate weighting we compute the frequencies of a weighted histogram. Either O_i and E_i must both be normalized, or, more simply and equivalently, the weighted frequencies $O_{i, w}$ must be scaled such that

$$\sum_{i=1}^{m} O_{i,w} = n, \quad (8)$$

where n is the total number of measurements in all m bins.

We implement the χ^2 test using `scipy`.[4]`stats.chisquare`, which evaluates the test statistic plus a p-value.

We apply the χ^2 test to both unweighted and weighted PA histograms, of both 2-axial and 4-axial PAs. For weighted histograms

the observed frequency $O_{i, w}$ is scaled (equation 8) and the expected frequency E_i is uniform and unweighted. In all cases the bin width is chosen to ensure $E_i > 5$.

3.2 Bimodality tests

As for uniformity tests, we perform bimodality tests on PAs after parallel transport to the centre of the A1 region (Hutsemékers 1998).

To examine bimodality we considered Hartigans' dip statistic (HDS; Hartigan & Hartigan 1985), the bimodality coefficient (BC; SAS Institute 2004), and Akaike's information criterion difference (AIC$_{diff}$; Akaike 1974). Freeman & Dale (2013) compare these measures, and report that HDS has the highest sensitivity, followed by BC, and that both methods are generally convergent. They found that AIC$_{diff}$ behaves quite differently, and erroneously identifies bimodality in their simulations and experimental data.

We note the bimodality coefficient is unsuitable for hypothesis significance testing and has an undesirable sensitivity to skew. We also found it inconsistent with different bin sizes and concluded it was too capricious for us to draw any conclusions from its results.

We therefore test the PA distribution for bimodality using Hartigans' dip statistic.

3.2.1 Hartigans' dip statistic

Hartigans' dip statistic (Hartigan & Hartigan 1985) is a non-parametric test of the unimodality of continuous data. A distribution is categorized as unimodal if its cumulative distribution function is convex up to its maximum gradient (which corresponds to the peak in the distribution) and concave afterwards, i.e. with a single inflection point. HDS quantifies how far the CDF departs from unimodality, and indicates the location(s) of any departure, i.e. the peak(s) in a bimodal (multimodal) distribution. This is well explained and illustrated by Maurus & Plant (2016).

We evaluate HDS using Benjamin Doran's PYTHON port of `unidip.UniDip`,[5] which follows Maurus & Plant (2016). The sensitivity of this test is controlled by the parameter α; we use $\alpha = 0.03$ to isolate peaks with at least 97 per cent signal-to-noise confidence. This implementation does not currently accommodate weighted data.

We evaluate Hartigans' dip statistic for continuous unweighted 2-axial PAs. We do not apply it to 4-axial PAs, since this conversion yields unimodal data, for which HDS is not meaningful.

3.3 Correlation tests

Uniformity and bimodality tests do not quantify the degree of alignment in the data; for this we need specific statistical methods appropriate to axial data on the celestial sphere. We considered two tests, the S test and the Z test, that have been widely used to analyse quasar-polarization alignments (e.g. Hutsemékers 1998; Jain et al. 2004; Pelgrims & Hutsemékers 2015), but are appropriate to analyse the alignment of any vectors on the celestial sphere (e.g. Contigiani et al. 2017).

Our sample of 71 LQG PAs is relatively small. Hutsemékers et al. (2014) report the Z test is better suited to small samples than the S test, because the latter uses a measure of angle dispersion which suffers reduced power with small samples. However, if PA alignment is 'global' (i.e. correlations are present throughout the survey area),

---

[3]Recommendation in electronic supplementary material 2.
[4]SciPy community project (Virtanen et al. 2020).

[5]https://github.com/BenjaminDoran/unidip

then PAs will not be correlated to positions and the power of the Z test will reduce dramatically (Pelgrims 2019, private communication).

We find that LQG PAs are not correlated with position, and that the S test is more appropriate than the Z test to quantify their alignment.

3.3.1 S test

The S test was developed by Hutsemékers (1998) and analyses the dispersion of vectors with respect to their nearest n_v neighbours (identified as explained in Appendix F1). For each vector i a measure of the dispersion d_i is calculated as

$$d_i(\theta) = 90 - \frac{1}{n_v} \sum_{k=1}^{n_v} |90 - |\theta_k - \theta||, \qquad (9)$$

where θ_k are the 2-axial PAs $[0°, 180°)$ of the neighbouring n_v vectors, including central vector i. The value of θ that minimizes the function $d_i(\theta)$ is a measure of the average PA at the location of i. Use of absolute values accounts for the axial nature of the data (Fisher 1993).

For vector i the mean dispersion D_i of its n_v nearest neighbours is calculated to be the minimum value of $d_i(\theta)$, which will be small for coherently aligned vectors. The measure of alignment within the whole sample of n vectors is given by the S test statistic

$$S_D = \frac{1}{n} \sum_{i=1}^{n} D_i, \qquad (10)$$

with one free parameter n_v. If the vectors are aligned, the value of S_D will be smaller than if they are uniformly distributed. So, the significance level for this version of the S test is evaluated as the probability that a random numerical simulation has a lower S_D than that observed (Cabanac et al. 2005).

Jain et al. (2004) introduce a coordinate invariant version of the S test, similar to the original except that, instead of the dispersion measure in equation (9), they use

$$d_i(\theta) = \frac{1}{n_v} \sum_{k=1}^{n_v} \cos[2\theta - 2(\theta_k + \Delta_{k \to i})], \qquad (11)$$

where $\Delta_{k \to i}$ is the angle by which the PA θ_k changes during parallel transport from position k to position i. Here, the factor two accounts for the axial nature of the data. The measure of dispersion is given by the *maximum* value of equation (11) (as opposed to the minimum value of equation 9). The S statistic is calculated as previously (equation 10). Pelgrims (2016) notes that the same value of θ that maximizes equation (11) at the same time minimizes equation (9), so the two versions are fully equivalent.

Jain et al. (2004) show the maximization of $d_i(\theta)$ is calculated analytically as

$$d_i\big|_{max} = \frac{1}{n_v} \left[\left(\sum_{k=1}^{n_v} \cos \theta_k' \right)^2 + \left(\sum_{k=1}^{n_v} \sin \theta_k' \right)^2 \right]^{1/2}, \qquad (12)$$

where $\theta_k' = 2(\theta_k + \Delta_{k \to i})$ is the circular version of θ_k after parallel transport to position i. We can similarly apply this to the 4-axial version of θ_k by using a factor of 4. This calculation is straightforward to code and avoids the time-consuming trials of the original version (Hutsemékers 1998). A large value of $d_i|_{max}$ indicates small dispersion, so a large value of S_D indicates strong alignment.

The significance level of the S test for alignment (orthogonality) is the probability that a random numerical simulation has a higher (lower) S_D than that observed. See Section F2 for an explanation of
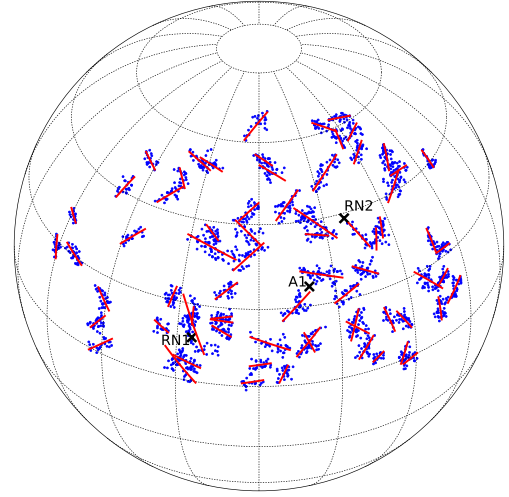

Figure 1. LQG quasars (blue dots) and ODR axes (red lines) shown on the celestial sphere (east to the right). Also shown, the centres (black crosses) of A1, RN1, and RN2 regions (Hutsemékers 1998; Pelgrims & Hutsemékers 2015). A1 is the parallel transport destination for uniformity and bimodality tests. Projection is centred on $\alpha = 180°$, $\delta = 35°$ (J2000), parallels and meridians are separated by $20°$. RA increases to the right.

this interpretation, and details of how we estimate significance level using numerical simulations.

We evaluate the S test for unweighted 2-axial and 4-axial PAs. For 2-axial we analyse PAs of the form $\Theta_{2ax} = 2\theta$, and for 4-axial we analyse PAs of the form $\Theta_{4ax} = 4\theta_{4ax}$ (see Appendix D). In both cases we apply parallel transport corrections before transforming the angles.

4 RESULTS: LQG POSITION ANGLES

We identify 71 LQGs of ≥ 20 quasars and detection significance $\geq 2.8\sigma$. The LQG positions on the celestial sphere, and their orientation as determined by the 2D method, are illustrated in Fig. 1. By eye it appears that the orientations may be somewhat preferentially aligned, but we caution that the orthographic projection may be deceiving. LQG positions in 3D proper space, and their orientations, as determined by the 3D method, are illustrated in Appendix G.

The results of both the 2D and 3D approaches are presented in Tables 1 (example LQGs) and C1 (full LQG sample, Appendix C2), and generally agree well. The PAs listed in these tables are measured *in situ* at the location of each LQG, and will have parallel transport corrections applied before statistical analysis (Section 5). For both approaches, bootstrap re-sampling with replacement is used to estimate the uncertainty in the form of the half-width confidence interval (HWCI, γ_h) of 10 000 bootstraps.

Figs 2 (histograms) and 3 (rose diagrams) show LQG PAs, after parallel transport, both unweighted and weighted by orthogonal distance regression goodness of fit (equation 1). For axial data $[0°, 180°)$, where $0°$ and $180°$ are equivalent, a conventional histogram (Fig. 2) can be misleading, since it represents data that are close together (e.g. $1°$ and $179°$) at opposite extremes of the distribution. An alternative representation is the rose diagram (Fig. 3), where wedge length is proportional to the count and spanning angle denotes the bins.

In both Figs 2 and 3 the data appear bimodal, with peaks at $\theta \sim 45°$ and $\theta \sim 135°$ (in the absence of goodness-of-fit weighting). The

Table 1. Example large quasar groups (LQGs), where m is the number of members, and $\bar{\alpha}$, $\bar{\delta}$, and $\bar{z}$ are the mean right ascension, declination, and redshift of the member quasars. The normalized goodness-of-fit weight w (equation 1) is scaled by $w_{71} = w \times 71$ for clarity, and to distinguish those LQGs weighted higher ($w_{71} > 1$) or lower ($w_{71} < 1$) than the mean $\bar{w}$. Position angle θ and half-width confidence interval γ_h are shown for both the 2D and 3D approaches. The ratio of LQG ellipsoid axes lengths (from the 3D approach, Appendix C) is given by a: b: c. See Appendix C2 for the full sample of 71 LQGs.

|	J2000 (°)				2D PA (°)		3D PA (°)		
m	$\bar{\alpha}$	$\bar{\delta}$	$\bar{z}$	w_{71}	θ	γ_h	θ	γ_h	a: b: c
20	121.1	27.9	1.73	1.13	119.7	10.4	115.1	10.7	0.50:0.30:0.21
20	151.5	48.6	1.46	1.21	144.4	9.2	144.3	11.5	0.50:0.33:0.17
20	155.9	12.8	1.50	0.32	120.7	25.2	117.2	37.0	0.44:0.43:0.14
20	163.6	16.9	1.57	1.00	0.9	8.8	5.7	9.2	0.47:0.33:0.20
					...				
23	209.5	34.3	1.65	1.99	152.5	2.8	152.3	2.8	0.68:0.17:0.15
23	214.3	31.8	1.48	0.27	18.6	32.6	87.7	32.5	0.47:0.35:0.18
					...				
26	160.3	53.5	1.18	0.33	110.6	23.2	111.5	25.1	0.38:0.34:0.28
26	171.7	24.2	1.10	0.78	48.5	8.3	47.3	8.1	0.46:0.29:0.24
					...				
55	196.5	27.1	1.59	0.95	107.5	3.4	107.0	3.4	0.58:0.24:0.18
56	167.0	33.8	1.11	0.81	110.2	3.5	110.4	3.8	0.50:0.29:0.21
64	196.4	39.9	1.14	0.83	133.6	3.0	133.9	3.2	0.48:0.36:0.17
73	164.1	14.1	1.27	0.76	156.6	4.2	156.3	4.5	0.55:0.28:0.16
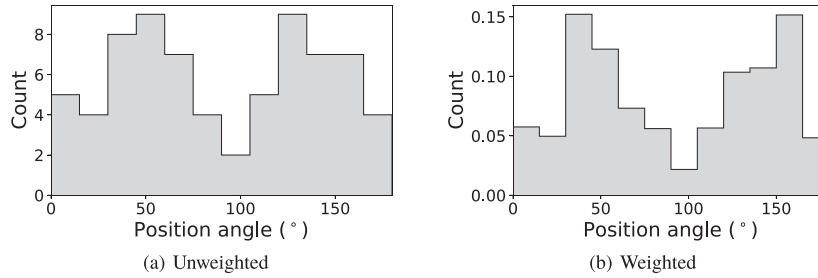
Figure 2. LQG position angles, all parallel transported to and measured at the centre of the A1 region (Hutsemékers 1998). (a) is unweighted and (b) is ODR goodness-of-fit weighted, both with 15° bins. The bimodal distribution is robust to whether or not ODR goodness-of-fit weighting is used.
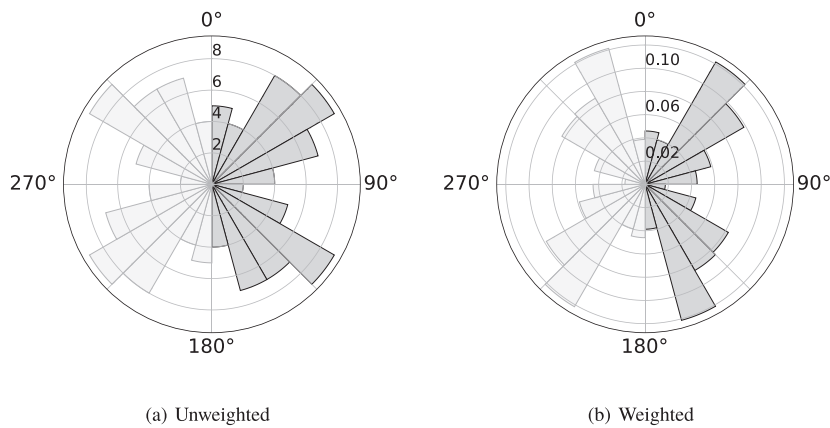

Figure 3. As Fig. 2 but represented as a rose diagram; again (a) is unweighted and (b) is ODR goodness-of-fit weighted, both with 15° bins. As is conventional for undirected axial data the [0°, 180°) data are duplicated on the opposite side of the rose diagram [180°, 360°). For clarity, we reiterate that the PA data are axial [0°, 180°) and not circular [0°, 360°); we do not have PAs in the range [180°, 360°) (the lighter shade).
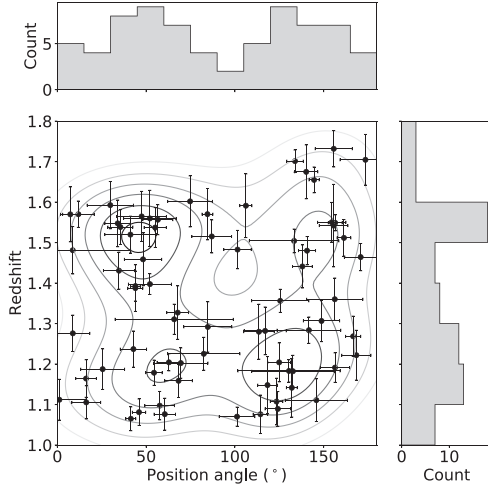
Figure 4. Marginal plot of the LQG PA × redshift plane. Error bars are PA half-width confidence interval and redshift standard deviation. Marginal histograms show PA and redshift distributions with $\Delta\theta = 15°$ and $\Delta z = 0.1$ bins. Contours are a Gaussian kernel density estimation of the scatter plot. Histograms and KDE are unweighted.

peaks are separated by $\Delta\theta \sim 90°$, indicating that some LQGs may have PAs that are preferentially parallel (i.e. aligned) while others are preferentially orthogonal to one another [this is described as 'anti-aligned' by Hutsemékers et al. (2014) and Pelgrims (2016)].

4.1 LQG PAs as a function of redshift

Fig. 4 shows a marginal plot of the LQG PA × redshift plane; redshift here is LQG redshift, defined as the mean redshift of its member quasars. From Fig. 2(a), we expect the PA histogram (top margin) to be bimodal, as seen. The redshift histogram (right margin) also exhibits some bimodality. To investigate the relationship between these two variables, and whether there is any correlation between their modes, we add kernel density estimation (KDE) contours to the scatter plot. This shows hints of three or four modes, although the correlation is weak. We note that the apparently stronger modes at $\theta \sim 45° \times z \sim 1.5$ and $\theta \sim 135° \times z \sim 1.2$ result from the points with the greatest uncertainty. Conversely, the weaker modes at $\theta \sim 55° \times z \sim 1.2$ and $\theta \sim 150° \times z \sim 1.5$ result from the points with the smallest uncertainty.

Due to their scale, LQGs extend considerably in the radial direction, and some features may be lost when we analyse only their mean redshift. Our sample of 71 LQGs collectively comprise 2076 member quasars. Fig. 5 shows the redshift distribution of these. The distribution appears bimodal with modes (peaks) at $z \sim 1.15$ and $z \sim 1.55$, similar to that for LQG redshifts (Fig. 4, right margin).

4.2 LQG PA weights

The 2D approach of determining position angles uses orthogonal distance regression (ODR) of tangent plane projected quasars. We evaluate the ODR goodness of fit (equation 1), which may be used to weight the PAs used for some of the statistical analysis (Section 5). An alternative empirical weighting scheme could use measurement uncertainties, or half-width confidence intervals (HW-CIs, Appendix E), estimated from 10 000 bootstraps.
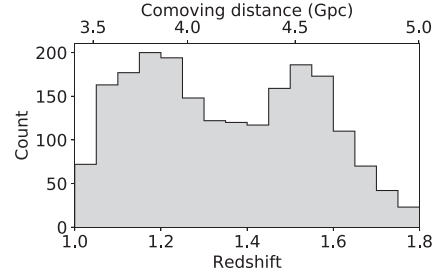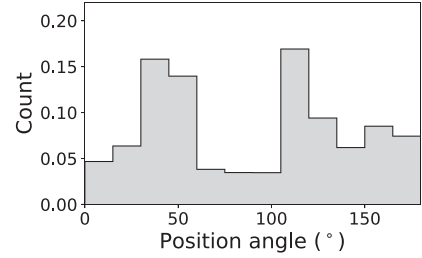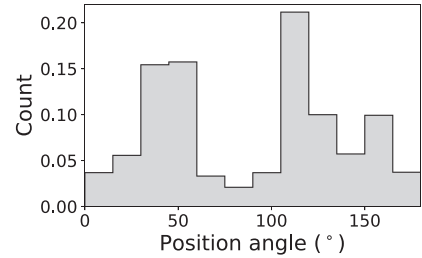

Figure 5. Redshift distribution of the 2076 member quasars comprising our sample of 71 LQGs, with $\Delta z = 0.05$ bins, unweighted. Top x-axis shows approximate proper radial distance. Distribution is bimodal, with modes (peaks) at $z \sim 1.15$ and $z \sim 1.55$.

(a) ODR GoF weighted

(b) HWCI weighted

Figure 6. Position angles calculated by 2D approach, with no parallel transport, and 15° bins. (a) is goodness-of-fit weighted and (b) is half-width confidence interval weighted. The bimodal distribution is robust to which method of weighting is used.

The PA distribution is robust between these alternative weighting schemes. Fig. 6 shows PAs, determined by the 2D approach, with no parallel transport, and weighted by (a) goodness-of-fit weights $w = \ell/\sigma^2$, and (b) HWCI weights $w = 1/\gamma_h$. The bimodal distribution is robust to which method of weighting is used. We continue to use the former in this work as it is more physically motivated and slightly more conservative.

5 RESULTS: STATISTICAL ANALYSIS

The position angle distribution of our sample of 71 LQGs appears bimodal, with modes at $\bar{\theta} \sim 52 \pm 2°$, $137 \pm 3°$ (with goodness-of-fit weighting) after parallel transport to the centre of the A1 region (Hutsemékers 1998). The median location of the peaks after parallel transport to all 71 LQG locations is $\bar{\theta} \sim 45 \pm 2°$, $136 \pm 2°$. The peaks are separated by $\Delta\theta \sim 90°$, indicating that some LQGs have

Table 2. Results (*p*-values) of uniformity tests. For 2-axial (4-axial) PAs the χ^2 test is evaluated using 20° (10°) bins. The χ^2 and Kuiper's tests are computed both with and without goodness-of-fit weighting. For 2-axial PAs, only the χ^2 test shows evidence for non-uniformity (of weighted PAs). For 4-axial PAs, all tests show evidence for non-uniformity, mostly marginal, with Kuiper's being the most significant.

| | | | *p*-value | |
| | 2-axial | | 4-axial | |
Test	Unweighted	Weighted	Unweighted	Weighted
χ^2	0.16	0.01	0.07	0.02
Kuiper's	0.62	0.59	0.009	0.008
HR	0.07	–	0.04	–

PAs that are preferentially aligned with each other, while others are preferentially orthogonal. We apply the statistical methods of Section 3 to analyse the uniformity, bimodality, and correlation of these PAs.

5.1 Uniformity: LQG PAs are unlikely to be uniform

The results from applying the uniformity tests are listed in Table 2. The χ^2 test does not show evidence of non-uniformity for 2-axial PAs without weighting (16 per cent significance level), but does show some evidence of non-uniformity with weighting (1 per cent significance level). For 4-axial PAs it shows marginal evidence of non-uniformity both with and without weighting (2 and 7 per cent, respectively).

Kuiper's test does not show evidence of non-uniformity for 2-axial PAs, with or without weighting. However, the PA distribution is bimodal, which dramatically reduces the test's discriminatory power, so the absence of a signal is unsurprising. For 4-axial PAs Kuiper's test indicates a rejection of the null hypothesis of uniformity at the 0.8 and 0.9 per cent significance level (weighted and unweighted), i.e. $\sim 2.4\sigma$.

Finally, the Hermans–Rasson test shows marginal evidence of non-uniformity for 2-axial and 4-axial PAs (7 and 4 per cent significance levels, respectively), both without weighting.

Based on the results from these three uniformity tests we cannot confidently reject the null hypothesis that the observed PAs are drawn from a uniform distribution. The most appropriate test for the axial and bimodal nature of the 2-axial PA data is the HR test, which shows marginal evidence of non-uniformity.

However, if we *a priori* expect *f*-fold symmetry (specifically 2-fold, in case of a bimodal distribution) then the conversion of PAs to 4-axial becomes physically well motivated as well as statistically legitimate. In this case, based on the 4-axial results of Kuiper's test, we could confidently reject the null hypothesis and conclude that the PAs are non-uniform.

5.1.1 Uniformity of mock LQG catalogues

The results from applying uniformity tests to the mock LQGs of Section 2.4 are listed in Table 3. The χ^2 and Hermans–Rasson tests show no evidence for non-uniformity of the mock LQGs. This result is consistent between 2 and 4-axial PAs, individual mocks, stacks of sub-volumes, and the stack of all 3296 mock LQGs.

Kuiper's test also shows no evidence for non-uniformity, except for the stack of all mock LQGs evaluated as 4-axial data (and then only marginally). This could be an artefact of the realizations not being truly independent, but if so it is unclear why this would reveal itself only in one of the six tests on this sample. Our concerns

Table 3. Results (*p*-values) of uniformity tests applied to mock LQGs. For 2-axial (4-axial) PAs the χ^2 test is evaluated using 15° (7.5°) bins, except for individual mocks where it is evaluated using 30° (15°) bins. For multiple samples the number of LQGs and *p*-values are means plus the standard error on the mean. All PAs are unweighted and are parallel transported to, and measured at, the centre of the A1 region (Hutsemékers 1998). Most samples and tests show no evidence for non-uniformity.

Sample(s)	No. of LQGs	test	*p*-value 2-axial	4-axial
110 individual mocks	30 ± 0.4	χ^2	0.53 ± 0.03	0.51 ± 0.03
		Kuiper's	0.53 ± 0.03	0.50 ± 0.03
		HR	0.48 ± 0.03	0.50 ± 0.03
10 stacks of 11 sub-volumes	330 ± 3	χ^2	0.47 ± 0.10	0.41 ± 0.08
		Kuiper's	0.56 ± 0.10	0.44 ± 0.09
		HR	0.46 ± 0.11	0.62 ± 0.09
1 stack of 110 mocks	3,296	χ^2	0.20	0.31
		Kuiper's	0.27	0.03
		HR	0.16	0.15

about the independence of this particular stack, the otherwise highly consistent results, and our caution about interpreting results manifest only in 4-axial data, led to this anomalous result being discredited. We therefore conclude that all three tests indicate statistical uniformity of the mock LQGs.

5.2 Bimodality: LQG PA distribution is bimodal

We calculate Hartigans' dip statistic (HDS) for continuous unweighted PA data and recover two peaks between $\sim 36° - 83°$ and $\sim 114° - 156°$, with 97 per cent confidence. The unweighted (weighted) means of PAs in these two ranges are $\sim 54 \pm 2°$ ($\sim 52 \pm 2°$) and $\sim 136 \pm 3°$ ($\sim 137 \pm 3°$), consistent with the peaks we see in the distribution of categorical PA data (e.g. Fig. 2). The bimodality is therefore unlikely to be an artefact of binning.

We apply HDS after all PAs are parallel transported to the centre of the A1 region. Recalling Section 2.3 (see also Jain et al. 2004), the process of parallel transport rotates PAs; the amount of rotation depending on the path taken (direction and distance). Therefore it is reasonable to check whether parallel transporting to a different location would affect the position of the PA peaks.

We parallel transport all 71 PAs to the location of each of the 71 LQGs, and at each one fit a double Gaussian to the unweighted histogram of these PAs. The height, width, and location of each Gaussian are fitted using PYTHON'S `scipy.optimize.leastsq`.

The location of the centres of the double Gaussian peaks, at each of the 71 LQG locations, are shown in Fig. 7. The median (mean) location of the first peak is $\sim 45 \pm 2°$ ($\sim 47 \pm 2°$), and of the second is $\sim 136 \pm 2°$ ($\sim 138 \pm 2°$). These are consistent (within the mean PA half-width confidence interval of $\sim 10°$) with the peaks identified by HDS at the centre of the A1 region ($\sim 54 \pm 2°$ and $\sim 136 \pm 3°$ for unweighted PAs).

Individual PA distributions at each of the 71 LQG locations are shown in Fig. H1 (Appendix H). At most ($\gtrsim 80$ per cent) the PAs show a similar bimodal distribution.

5.3 Correlation: LQG PAs are aligned and orthogonal

We compute the S statistic S_D using the Jain et al. (2004) coordinate invariant version of the S test for samples of n_v nearest neighbours, where $10 \leq n_v \leq 70$. In Fig. 8, we show the values of S_D calculated for observed LQG PAs, represented as both 2-axial (green circles) and
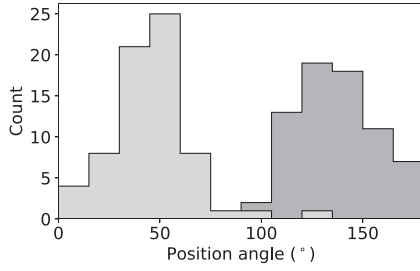
Figure 7. Location of the centres of the double Gaussian peaks in Fig. H1; 15° bins. The locations of the two peaks are consistent, regardless of parallel transport destination. The mean locations are ∼47 ± 2° and ∼138 ± 2°, consistent (within ≲10°) with the HDS result.
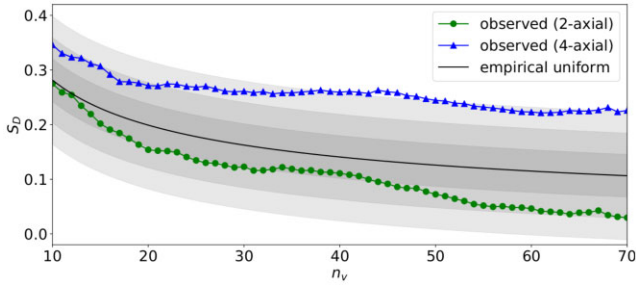

Figure 8. The S test statistic S_D calculated for 2-axial (green circles) and 4-axial (blue triangles) PAs as a function of nearest neighbours n_v determined in 3D. Also shown, empirical values estimated for a uniform distribution (black) and their approximate ±1σ, 2σ, 3σ confidence intervals (grey). 4-axial PAs show more alignment than uniform, while 2-axial show less.

4-axial (blue triangles) data, as a function of nearest neighbours n_v. Nearest neighbours are determined in 3D, taking into account radial distance. Results are entirely consistent with neighbours determined in 2D (i.e. by angular separation).

Larger values of S_D indicate stronger alignment. So Fig. 8 indicates that 4-axial LQG PAs are more aligned than expected if they are randomly drawn from a uniform distribution. Conversely, it also suggests that 2-axial LQG PAs are less aligned (or more orthogonal) than expected. The physical interpretation of the latter is not straightforward, but likely to be due to the contribution from orthogonal PAs leading to large dispersion, and hence a small value of S_D.

The approximate empirical standard deviation, and hence the ±1σ, 2σ, and 3σ confidence intervals shown in Fig. 8, are valid only for large n while $n_v \ll n$ (Jain et al. 2004). These are not valid assumptions for much of our range of S_D. Therefore, due to this and the mutual dependence between groups of nearest neighbours, we calculate the significance level of the S test using numerical simulations. In Fig. 9, we show the significance level of the S test calculated using 10 000 numerical simulations, for values of S_D determined using both 2-axial (green circles) and 4-axial (blue triangles) PAs. Again, this is shown as a function of nearest neighbours n_v.

With 4-axial LQG PAs we test for alignment and orthogonality by combining the modes, resulting in an alignment only signal (right-hand S_D tail). We find significance levels generally between 1 and 5 per cent for most numbers of nearest neighbours n_v. For most of the n_v range ($30 \leq n_v \leq 70$) the significance level is 0.9 per cent ≤ SL ≤
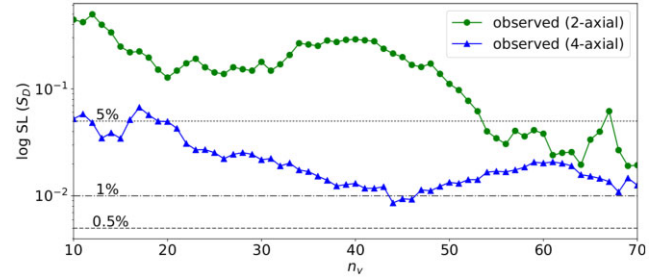

Figure 9. The logarithmic significance level (SL) of the S test calculated for 2-axial (green circles) and 4-axial (blue triangles) PAs as a function of nearest neighbours n_v determined in 3D. The dotted, dash–dotted, and dashed horizontal lines indicate SL = 0.05, 0.01, and 0.005, respectively. 2-axial LQG PAs show correlation ($\overline{SL} \sim 3.3$ per cent) above $n_v \geq 54$. 4-axial PAs show correlation ($\overline{SL} \sim 1.5$ per cent) above $n_v \geq 30$.

2.2 per cent, with a mean (median) of 1.5 per cent (1.5 per cent). It is most significant for $n_v \sim 45$, with SL ≃ 0.8 per cent.

For 2-axial LQG PAs the orthogonal mode appears to dominate (left-hand S_D tail). We find significance levels generally above 5 per cent until $n_v \geq 54$. For the remainder of the range of n_v ($54 \leq n_v \leq 70$) the SL is 1.9 per cent ≤ SL ≤ 6.2 per cent, with a mean (median) of 3.3 per cent (3.4 per cent).

5.3.1 Typical angular and proper separations

The number of LQG nearest neighbours n_v is a free parameter explored by the S test. We identify these neighbours using the 3D proper positions of each LQG centroid. The parameter n_v is related to the scale of the nearest neighbour groups, but because LQGs are not homogeneously distributed we cannot directly interpret it as corresponding to a particular scale. We evaluate the relationship between the parameter n_v and the typical scale of the nearest neighbour groups, defined as the median separation between each LQG and its n_v nearest neighbours.

Using these relationships, we evaluate the S test significance levels as a function of typical separation instead of nearest neighbours n_v. The functions do not differ significantly in shape, because the relationships are generally linear when $n_v \gtrsim 20$. We find that the correlation is most significant for typical angular (proper) separations of ∼30° (1.6 Gpc).

6 DISCUSSION AND CONCLUSIONS

We find that LQG PAs are unlikely to be drawn from a uniform distribution (p-values $0.008 \lesssim p \lesssim 0.07$). However, similar non-uniformity is not found in mock LQG catalogues, indicating the LQG correlation is not found in cosmological simulations. Further, the LQG PA distribution is bimodal, with modes for weighted PAs at $\bar{\theta} \sim 52 \pm 2°$, $137 \pm 3°$ (97 per cent confidence). This bimodality is robust to parallel transport destination, with the median location of the peaks at all 71 LQG locations of $\bar{\theta} \sim 45 \pm 2°$, $136 \pm 2°$. These angles are remarkably close to the mean angles of radio quasar polarization of $\bar{\theta} \simeq 42°$ and $\bar{\theta} \simeq 131°$, reported by Pelgrims & Hutsemékers (2015)[6] in two regions coincident with our LQG sample.

LQGs are aligned and orthogonal across very large scales, with a maximum significance of $\simeq 0.8$ per cent (2.4σ) for groups of ∼45 nearest neighbours, corresponding to typical angular (proper)

[6]Errors on these means were not reported.

separations of $\sim 30°$ (1.6 Gpc). The statistical significance of this correlation is marginal, therefore we cannot exclude it being a chance statistical anomaly. However, its coincidence with regions of quasar-polarization alignment (e.g. Hutsemékers 1998; Pelgrims & Hutsemékers 2015), the link between quasar polarization and LQG axes (Hutsemékers et al. 2014; Pelgrims & Hutsemékers 2016), and the similarity between LQG position angles and the preferred angles of quasar radio polarization alignment (Pelgrims & Hutsemékers 2015), suggest an interesting result.

We find no indication that boundary effects or selection effects have influenced these results. Three of the LQGs might be truncated by the RA, Dec. boundaries of DR7QSO: removing them from consideration made no significant difference. Randomly generated LQGs with related parameters (same encompassing circle, same number of members) to the real LQGs did not reproduce the results.

A plausible mechanism for the correlation of LQG orientations on such large scales is not obvious. We considered the geometry of the Universe. van de Weygaert (2007) uses Voronoi tessellation (Voronoï 1908) to describe the observed cosmic web on $\gtrsim 100$ Mpc scales (see also Icke & van de Weygaert 1987; van de Weygaert 1994). We speculated whether a cellular structure to the Universe, such as Voronoi tessellation, or a more regular crystalline structure, could cause such an effect.

We also considered primordial anisotropies. Poltis & Stojkovic (2010) proposed cosmic strings as an explanation for quasar-polarization alignments. They suggest the decay of these would seed correlated primordial magnetic fields. However, using the CMB, the possible amplitude of these has been constrained to less than a few nanoGauss (Planck Collaboration et al. 2016). Hutsemékers et al. (2005) suggest the apparent rotation of mean optical quasar polarization angle with redshift may be caused by a global rotation of the Universe, such as that invoked by Jaffe et al. (2005) to explain large-scale anisotropies in the CMB data. However, from CMB temperature and polarization analysis Saadeh et al. (2016) conclude that the Universe is neither rotating nor anisotropically stretched.

For the geometric and primordial explanations we considered, it is unclear how they could translate into our observed position angle distribution. Further, the primordial explanations have been disfavoured by observations. The origin of the LQG orientation correlation remains unexplained.

We found no evidence of ΛCDM cosmological simulations predicting correlations between objects on Gpc scales, but this had not been specifically examined for LQGs. Using mock LQG catalogues (Marinello et al. 2016) we found no evidence of LQG correlation in the Horizon Run 2 simulation (Kim et al. 2011). This suggests that the cosmic web of the observed Universe differs on the largest scales to this dark-matter-only *N*-body simulation. It hints that there could, given the caveats associated with the simulations, be aspects of the large-scale structure that are not captured by the power spectrum. Running the LQG finder on other cosmological simulations would be informative. If the correlation in LQG orientation is confirmed then perhaps it is an unexpected feature of known physics. If it is not seen then perhaps something is missing from the simulations (e.g. primordial anisotropies) or it is, after all, a statistical fluke which coincidentally gives rise to the aligned quasar polarizations.

The LQG orientation correlation we found offers a plausible explanation for the quasar-polarization alignments reported by many studies (e.g. Hutsemékers 1998; Hutsemékers & Lamy 2001; Jain et al. 2004; Cabanac et al. 2005; Tiwari & Jain 2013; Pelgrims & Cudell 2014; Pelgrims & Hutsemékers 2015; Pelgrims 2019). If LQG axes are preferentially aligned at $\bar{\theta} \sim 45 \pm 2°$, $136 \pm 2°$ (this work, median modes), and if quasar polarization vectors are preferentially

parallel and orthogonal to LQG axes (Hutsemékers et al. 2014; Pelgrims & Hutsemékers 2016), this could result in polarization vectors with preferred angles of $\sim 42°$ and $\sim 131°$ (Pelgrims & Hutsemékers 2015). Our results therefore offer corroborating evidence for, and enhancement of, the intrinsic alignment interpretation of these studies.

Quasar-polarization alignment is also detected in the south Galactic cap (SGC; e.g. Hutsemékers 1998; Pelgrims & Hutsemékers 2015), which is not coincident with our LQG sample in the north Galactic cap (NGC). The forthcoming 4-m Multi-Object Spectroscopic Telescope (4MOST) Active Galactic Nuclei survey (Merloni et al. 2019) will survey a million $z \lesssim 2.5$ quasars over $\sim 10\,000\,\mathrm{deg}^2$, with first light expected in 2022. This could deliver an LQG sample in the SGC for similar evaluation to our work in the NGC. Of particular interest would be whether LQG orientation again corresponds to the preferred angle of quasar radio polarization alignment, which differ between NGC and SGC (e.g. Pelgrims 2016).

Our results are based on a sample of 71 LQGs at redshifts $1.0 \leq z \leq 1.8$, which were detected using the SDSS DR7QSO catalogue (Schneider et al. 2010) of ~ 105k quasars across $\sim 7600\,\mathrm{deg}^2$. Forthcoming spectroscopic surveys will deliver a far larger sample of quasars, e.g. the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016) 5-yr survey aims to target 1.7 million $z < 2.1$ quasars covering $\sim 14\,000\,\mathrm{deg}^2$, beginning in May 2021. This has the potential to deliver a larger sample of LQGs for a better assessment of their correlation.

If the LQG orientation correlation is real, it represents large-scale structure alignment over $\gtrsim$Gpc scales, larger than those predicted by cosmological simulations and at least an order of magnitude larger than any so far observed, with the exception of quasar-polarization/radio-jet alignment. Careful statistical analysis is required before making inferences about whether such a large-scale correlation challenges the assumption of large-scale statistical isotropy and homogeneity of the Universe.

To conclude, we find large-scale correlation of LQG orientations, which we report here for the first time. This helps explain a substantial body of work on quasar-polarization/radio-jet alignment, but at the expense of raising potentially even more challenging questions about the origin of the LQG correlation and its implications for isotropy and homogeneity. Forthcoming surveys and the other future work we suggest here will illuminate LQGs and their intriguing correlation further.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The datasets were derived from sources in the public domain: https://classic.sdss.org/dr7/products/value_added/qsocat_dr7.html.

REFERENCES

Agarwal N., Kamal A., Jain P., 2011, Phys. Rev. D, 83, 065014
Akaike H., 1974, IEEE Trans. Autom. Control, 19, 716
Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17
Baugh C. M. et al., 2019, MNRAS, 483, 4922
Blinov D., Casadio C., Mandarakas N., Angelakis E., 2020, A&A, 635, A102

Cabanac R. A., Hutsemékers D., Sluse D., Lamy H., 2005, in Adamson A., Aspin C., Davis C., Fujiyoshi T., eds, ASP Conf. Ser. Vol. 343, Astronomical Polarimetry: Current Status and Future Directions. Astron. Soc. Pac., San Francisco, p. 498.

Challinor A., Chon G., 2002, Phys. Rev. D, 66, 127301

Clowes R. G., Campusano L. E., Graham M. J., Söchting I. K., 2012, MNRAS, 419, 556

Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Söchting I. K., Graham M. J., 2013, MNRAS, 429, 2910

Codis S., Jindal A., Chisari N. E., Vibert D., Dubois Y., Pichon C., Devriendt J., 2018, MNRAS, 481, 4753

Contigiani O. et al., 2017, MNRAS, 472, 636

Das S., Jain P., Ralston J. P., Saha R., 2005, J. Cosmol. Astropart. Phys., 2005, 002

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

DESI Collaboration, 2016, preprint (arXiv:1611.00036)

Dubois Y. et al., 2014, MNRAS, 444, 1453

Efron B., 1979, Ann. Statist., 7, 1

Einasto M., Tago E., Jaaniste J., Einasto J., Andernach H., 1997, A&AS, 123, 119

Einasto M. et al., 2014, A&A, 568, A46

Feigelson E. D., Babu G. J., 2012, Modern Statistical Methods for Astronomy. Cambridge Univ. Press, Cambridge

Fisher N. I., 1993, Statistical Analysis of Circular Data. Cambridge Univ. Press, Cambridge

Freeman J. B., Dale R., 2013, Behav. Res., 45, 83

Ganeshaiah Veena P., Cautun M., van de Weygaert R., Tempel E., Jones B. J. T., Rieder S., Frenk C. S., 2018, MNRAS, 481, 414

Graham M. J., Clowes R. G., Campusano L. E., 1995, MNRAS, 275, 790

Hartigan J. A., Hartigan P. M., 1985, Ann. Statist., 13, 70

Hermans M., Rasson J. P., 1985, Biometrika, 72, 698

Hewett P. C., Wild V., 2010, MNRAS, 405, 2302

Hutsemékers D., 1998, A&A, 332, 410

Hutsemékers D., Lamy H., 2001, A&A, 367, 381

Hutsemékers D., Cabanac R., Lamy H., Sluse D., 2005, A&A, 441, 915

Hutsemékers D., Borguet B., Sluse D., Cabanac R., Lamy H., 2010, A&A, 520, L7

Hutsemékers D., Payez A., Cabanac R., Lamy H., Sluse D., Borguet B., Cudell J., 2011, in Bastien P., Manset N., Clemens D. P., St-Louis N., eds, ASP Conf. Ser. Vol. 449, Astronomical Polarimetry 2008: Science from Small to Large Telescopes. Astron. Soc. Pac., San Francisco, p. 441

Hutsemékers D., Braibant L., Pelgrims V., Sluse D., 2014, A&A, 572, A18

Icke V., van de Weygaert R., 1987, A&A, 184, 16

Isobe T., Feigelson E. D., Akritas M. G., Babu G. J., 1990, ApJ, 364, 104

Jackson J. C., 1972, MNRAS, 156, 1P

Jaffe T. R., Banday A. J., Eriksen H. K., Górski K. M., Hansen F. K., 2005, ApJ, 629, L1

Jain P., Narain G., Sarala S., 2004, MNRAS, 347, 394

Joshi S. A., Battye R. A., Browne I. W. A., Jackson N., Muxlow T. W. B., Wilkinson P. N., 2007, MNRAS, 380, 162

Kaiser N., 1987, MNRAS, 227, 1

Kim J., Park C., Rossi G., Lee S. M., Gott J. R. III., 2011, JKAS, 44, 217

Kraljic K., Davé R., Pichon C., 2020, MNRAS, 493, 362

Kuiper N. H., 1960, Indag. Math., 63, 38

Landler L., Ruxton G. D., Malkemper E. P., 2018, Behav. Ecol. Sociobiol., 72, 128

Landler L., Ruxton G. D., Malkemper E. P., 2019, BMC Ecol., 19, 30

Mandarakas N., Blinov D., Casadio C., Pelgrims V., Kiehlmann S., Pavlidou V., Tassis K., 2021, A&A, 653, A123

Mardia K., Jupp P., 2000, Directional Statistics. Wiley Series in Probability and Statistics, Wiley

Marinello G. E., 2015, PhD thesis, University of Central Lancashire

Marinello G. E., Clowes R. G., Campusano L. E., Williger G. M., Söchting I. K., Graham M. J., 2016, MNRAS, 461, 2267

Maurus S., Plant C., 2016, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. Association for Computing Machinery, New York, p. 1055

Merloni A. et al., 2019, The Messenger, 175, 42

Monahan J. F., 2011, Numerical Methods of Statistics, 2 edn. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, Cambridge

Nadathur S., 2013, MNRAS, 434, 398

Park C., Choi Y.-Y., Kim J., Gott J. R. III, Kim S. S., Kim K.-S., 2012, ApJ, 759, L7

Park C., Song H., Einasto M., Lietzen H., Heinamaki P., 2015, JKAS, 48, 75

Payez A., Cudell J. R., Hutsemékers D., 2008, in Cugon J., Lansberg J.-P., Matagne N., eds, AIP Conf. Ser. Vol. 1038, Hadronic Physics: Joint Meeting Heidelberg-Liège-Paris-Wroclaw - HLPW 2008. Am. Inst. Phys., New York, p. 211

Payez A., Cudell J. R., Hutsemékers D., 2011, Phys. Rev. D, 84, 085029

Pelgrims V., 2016, PhD thesis, Université de Liège

Pelgrims V., 2019, A&A, 622, A145

Pelgrims V., Cudell J. R., 2014, MNRAS, 442, 1239

Pelgrims V., Hutsemékers D., 2015, MNRAS, 450, 4161

Pelgrims V., Hutsemékers D., 2016, A&A, 590, A53

Pen U.-L., Lee J., Seljak U., 2000, ApJ, 543, L107

Pereyra L. A., Sgró M. A., Merchán M. E., Stasyszyn F. A., Paz D. J., 2020, MNRAS, 499, 4876

Pilipenko S., 2007, Astr. Rep., 51, 820

Pilipenko S., Malinovsky A., 2013, preprint (arXiv:1306.3970)

Planck Collaboration XIX, 2016, A&A, 594, A19

Poltis R., Stojkovic D., 2010, Phys. Rev. Lett., 105, 161301

Press W. H., Davis M., 1982, ApJ, 259, 449

Richards G. T. et al., 2006, AJ, 131, 2766

Saadeh D., Feeney S. M., Pontzen A., Peiris H. V., McEwen J. D., 2016, Phys. Rev. Lett., 117, 131302

SAS Institute, 2004, SAS/STAT 9.1 User's Guide

Schneider D. P. et al., 2010, AJ, 139, 2360

Taylor A. R., Jagannathan P., 2016, MNRAS, 459, L36

Tempel E., Tamm A., 2015, A&A, 576, L5

Tiwari P., Jain P., 2013, Int. J. Mod. Phys. D, 22, 1350089

Tiwari P., Jain P., 2019, A&A, 622, A113

van de Weygaert R., 1994, A&A, 283, 361

van de Weygaert R., 2007, preprint (arXiv:0707.2877)

Vanden Berk D. E. et al., 2005, AJ, 129, 2047

Virtanen P. et al., 2020, Nat Methods, 17, 261

Voronoï G., 1908, J. für die Reine und Angew., 134, 198

Welker C. et al., 2020, MNRAS, 491, 2864

York D. G. et al., 2000, AJ, 120, 1579

Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

APPENDIX A: LQG FINDER

Clowes et al. (2012, 2013) detect LQG candidates using a 3D single-linkage hierarchical clustering algorithm, also known as friends-of-friends (FoF). This is equivalent to a 3D minimal spanning tree (MST). These type of methods are widely used to detect galaxy clusters, superclusters, voids, and filaments (e.g. Press & Davis 1982; Einasto et al. 1997; Park et al. 2012; Pereyra et al. 2020), as well as LQGs (e.g. Clowes et al. 2012, 2013; Nadathur 2013; Einasto et al. 2014; Park et al. 2015). They make no assumptions about cluster morphology.

Using the terminology of Barrow, Bhavsar & Sonoda (1985), MST treats the data set as a graph made up of vertices (nodes, in this case quasars) which are connected by edges (straight lines). This is a 'tree' when it has no closed paths (sequence of edges) and a 'spanning' tree when it contains all the vertices. For any graph, there are multiple possible spanning trees; the 'minimal' spanning tree is that of minimal length (sum of edge lengths). To identify clusters within the MST it may be separated, where edges exceeding a certain length are removed, leaving groups of objects with mutual separations less than this 'linkage length'.

The choice of linkage length is crucial – too long and clusters merge to fill the entire volume, too short they breakup into pairs and triplets (Graham, Clowes & Campusano 1995). Pilipenko (2007) categorizes the criteria for making this choice as physical or formal. With a priori knowledge of physical parameters (e.g. size, membership, and density) of the clusters, it is possible to choose the scale that maximizes the fraction of clusters with those parameters. The criterion used by Graham et al. (1995) is an example of a physical approach; they choose the scale that maximizes the number of clusters of a minimum membership. An example formal approach would be to choose a scale based on the mean nearest-neighbour separation.

Clowes et al. (2012, 2013) use this latter approach. Their quasar sample has a mean nearest-neighbour separation of ∼74 Mpc. They also account for uncertainties in their edge lengths (i.e. proper distances) due to redshift errors and peculiar velocities (for estimates, see Clowes et al. 2012) and choose a linkage length of 100 Mpc.

Clowes et al. (2012, 2013) estimate the overdensity and statistical significance of their LQGs using a convex hull of member spheres (CHMS) method. This is described in detail in Clowes et al. (2012), but briefly, for each LQG they calculate the volume of a convex hull of spheres of radius half the mean edge length at each vertex (member quasar location). The CHMS volume of an LQG of m members is then compared with the distribution of CHMS volumes of clusters of m points in Monte Carlo simulations of the same size and density as their control area, in order to estimate overdensity and statistical significance. We note this method of estimating statistical significance is not universally accepted (Nadathur 2013; Pilipenko & Malinovsky 2013; Park et al. 2015).

The Huge-LQG and Clowes-Campusano LQG, amongst others, have been independently detected using different FoF algorithms (Nadathur 2013; Einasto et al. 2014; Park et al. 2015). Indeed, and unsurprisingly, our LQG sample has many objects in common with the publicly available catalogue[7] of Einasto et al. (2014). Einasto et al. (2014) appear to have followed closely the approach of Clowes et al. (2012, 2013) in terms of input data, selection algorithm, linkage scale, and cosmological model and parameters, but applied to a reduced area of DR7QSO. Einasto et al. (2014) differed substantially, however, in not providing any measures of statistical significance or overdensity, which are important for ranking the LQG candidates.

[7] https://vizier.u-strasbg.fr/viz-bin/VizieR?-source = J/A + A/568/A46

APPENDIX B: LQG SAMPLE

Fig. B1 illustrates the 71 LQGs in our sample, shown in tangent plane projection.
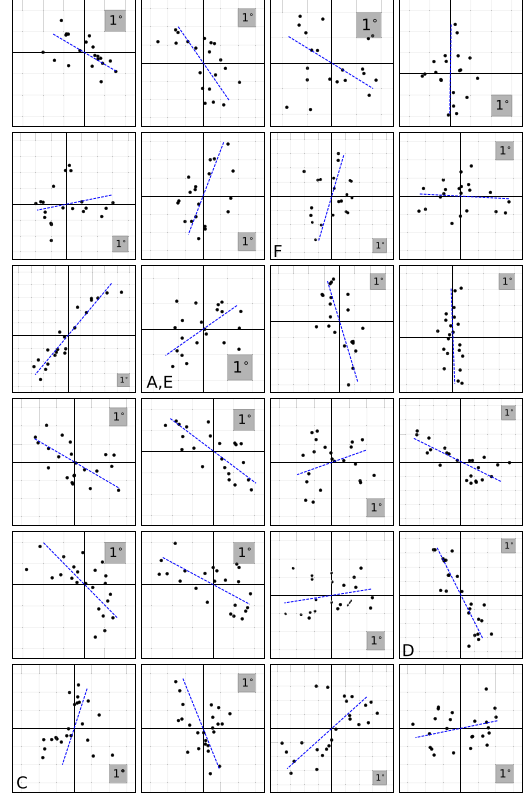

Figure B1. (a) Tangent plane projection of 1–24 (across then down) of the 71 LQGs in our sample, in Cartesian coordinates. Member quasars shown as black dots, orthogonal distance regression fit shown as dashed blue line. Solid black lines indicate $x = 0$ and $y = 0$, and grey square illustrates scale (1°). LQGs labelled A, C (D, E, F) are discussed in Fig. C3 (Fig. E1).
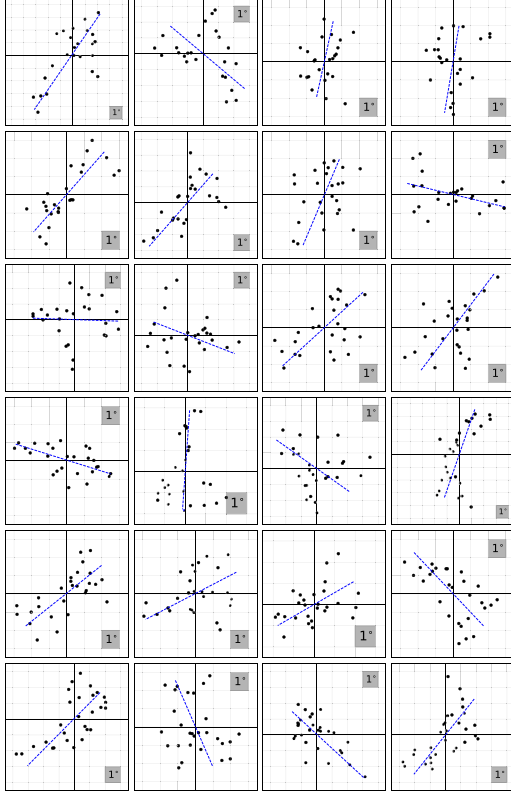
Figure B1. (b) Tangent plane projection of 25–48 (across then down) of the 71 LQGs in our sample, in Cartesian coordinates. Other plot details are as for Fig. B1 panel (a).
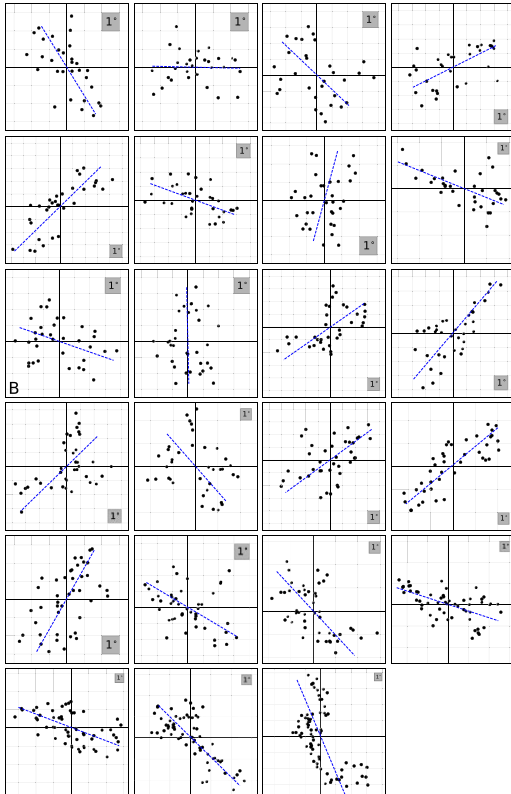

Figure B1. (c) Tangent plane projection of 49–71 (across then down) of the 71 LQGs in our sample, in Cartesian coordinates. Other plot details are as for Fig. B1 panel (a). LQG labelled B is discussed in Fig. C3.

APPENDIX C: LQG ORIENTATION

C1 LQG position angles – presentation

The position angle (PA) of a large quasar group can be calculated in either two or three dimensions. In two dimensions we treat the quasars as points on the celestial sphere, whereas in three dimensions we take into account their proper radial distances.

For the 2D approach quasar positions (in right ascension and declination) are projected on to the tangent plane as Cartesian (x, y) points. This plane meets the celestial sphere at the centre of gravity of the LQG $(x, y) = (0, 0)$, calculated assuming quasars are point-like unit masses. To determine LQG orientation we use orthogonal distance regression (ODR) of these projected points (see Fig. C1 for an example). This minimizes the sum of the squares of the orthogonal residuals between the points and the line (for a discussion of OLS, ODR, and other regression methods, see Isobe et al. 1990).

For the 3D approach, the covariance matrix of the quasar proper positions is decomposed into its eigenvectors and eigenvalues. Again, quasars are assumed to be point-like unit masses. The axes of a confidence ellipsoid (e.g. Fig. C2) are constructed from the eigenvectors and eigenvalues; each axis is in the direction of its eigenvector and its length ℓ is a function of its eigenvalue λ, specifically $\ell \propto \sqrt{\lambda}$. The first principal component (ellipsoid major axis) is given by the eigenvector with the largest eigenvalue. Finally, this is projected on to the plane orthogonal to the line-of-sight to define the (2D projected) PA.

PAs determined by the two approaches may differ, for example the 2D approach may be susceptible to projection effects and the 3D approach may be susceptible to redshift-space distortions. The orientation of the LQG with respect to the line of sight and its morphology may also induce differences. We expect PAs determined by the two approaches to agree well when the LQG is linear and orthogonal to the line of sight, but they may differ significantly when the LQG is broad, crooked, curved or aligned along the line of sight.

For our sample of 71 LQGs we find that the two approaches are generally consistent. Fig. C3 shows the PAs calculated using both the 2D and 3D approaches. Note that the PAs have not been parallel transported, but these angles serve as a useful comparison between the two approaches. The error bars are the half-width confidence
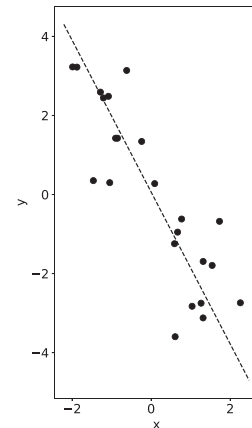

Figure C1. Tangent plane projection of an LQG, with coordinates (α, δ) projected to (x, y). Member quasars are shown as black dots and orthogonal distance regression fit shown as dashed line. Axes are labelled in degrees. This LQG is also shown in Figs B1 and E1, labelled D. Using the 2D PA approach, PA = 152.5°.
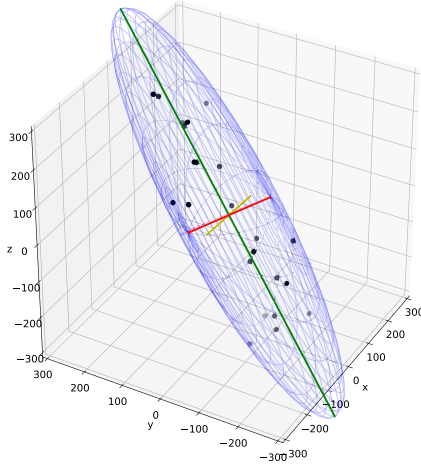
Figure C2. Same LQG as Fig. C1, but shown in 3D proper coordinates oriented with line-of-sight orthogonal to the page. Axes of enclosing ellipsoid (green = a-axis, yellow = b-axis, red = c-axis) constructed from eigenvectors and eigenvalues. Axes are labelled in Mpc. Using the 3D PA approach, 2D projected PA = $152.3°$.
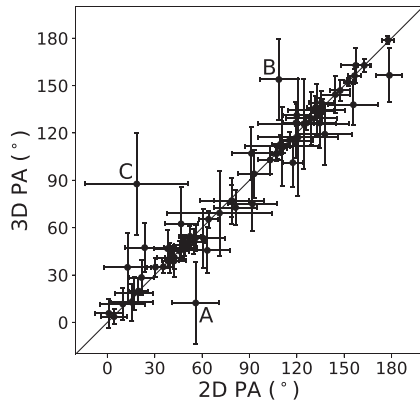

Figure C3. Position angles of 71 LQGs calculated by 2D and 3D approaches. The two generally agree well, with the three widest outliers (A, B, and C) due to the geometry of those particular LQGs (see text and Fig. B1). Error bars show half-width confidence intervals estimated using bootstrap re-sampling.

intervals estimated using bootstrap re-sampling. Note that, usually, the measurement uncertainties are slightly larger for the 3D approach.

The three widest outliers from the 1:1 diagonal line in Fig. C3 are due to the geometry of these particular LQGs (A, B, C; also labelled in Fig. B1). Two of these LQGs (A and B) have their major axes oriented towards the line of sight, and not significantly

longer than their first minor axes. Indeed, for both of these, the 2D approach fits a regression comparable to the first minor axes rather than the major axes. One LQG (C) is very irregular so linear fits are poor, and corresponding PAs are uncertain, in both the 2D and 3D approaches. The PAs of all three of these LQGs are given little weight by goodness-of-fit weighting (Section 2.2).

Both the 2D and 3D approaches have been used to determine the PAs of LQGs. Hutsemékers et al. (2014) use the 2D approach to demonstrate alignment of quasars' optical linear polarization with LQG axes, while Pelgrims & Hutsemékers (2016) use the 3D approach to evidence alignment of quasars' radio polarization with more LQG axes. The latter derived eigenvectors and eigenvalues from the inertia tensor rather than covariance matrix; results are equivalent. Pelgrims & Hutsemékers (2016) report that for the 2D approach PAs calculated using ODR are consistent (within $1°$) with those determined using the inertia tensor.

Pelgrims & Hutsemékers (2016) show that both approaches usually agree well, and argue that the 3D approach is more physically motivated. We agree, but note that any 3D analysis is susceptible to redshift errors. The quasars in our sample are from SDSS DR7QSO, which typically has quoted redshift errors of $\Delta z \sim 0.004$ (Schneider et al. 2010). There is also evidence for systematic errors of $\Delta z \sim 0.003$ (Hewett & Wild 2010). Using Monte Carlo simulations we find that these redshift errors introduce uncertainty in the PA, generally of a few degrees, but up to $\sim 30°$ for LQGs particularly oriented along the line of sight (e.g. LQGs A and B).

Furthermore, in the 3D approach, there is also the potential for errors due to redshift-space distortions from the quasars' peculiar velocities, causing their real-space distribution to be either elongated (Jackson 1972) or squashed (Kaiser 1987) along the line of sight. We find that the measurement uncertainties (Appendix E) are slightly larger for the 3D approach. We therefore base our analysis on the 2D approach; tangent plane projection of the LQG and orthogonal distance regression of the projected quasars.

C2 LQG position angles – tabulation

Table C1 presents the results of both the 2D and 3D approaches to determining position angle. To recap, the 2D approach involves tangent plane projection of the LQG quasars, followed by orthogonal distance regression (ODR) of the projected points. The 3D approach requires determining the proper coordinates of the LQG quasars, performing principal component analysis on the covariance matrix of these, then tangent plane projection of the resultant major axis.

The results of both approaches generally agree well. For both approaches, bootstrap re-sampling with replacement is used to estimate the uncertainty in the form of the half-width confidence interval (HWCI, γ_h) of 10 000 bootstraps. The PAs listed in Table C1 are measured *in situ* at the location of each LQG, and will have parallel transport corrections applied before statistical analysis.

Table C1. The 71 large quasar groups, where m is the number of members, and $\bar{\alpha}$, $\bar{\delta}$, and $\bar{z}$ are the mean right ascension, declination, and redshift of the member quasars. The normalized goodness-of-fit weight w (equation 1) is scaled by $w_{71} = w \times 71$ for clarity, and to distinguish those LQGs weighted higher ($w_{71} > 1$) or lower ($w_{71} < 1$) than the mean $\bar{w}$. Position angle θ and half-width confidence interval γ_h are shown for both the 2D and 3D approaches. The ratio of LQG ellipsoid axes lengths (from the 3D approach, Appendix C) is given by a: b: c. This list was summarized in Table 1.

	J2000 (°)				2D PA (°)		3D PA (°)		
m	$\bar{\alpha}$	$\bar{\delta}$	$\bar{z}$	w_{71}	θ	γ_h	θ	γ_h	a: b: c
20	121.1	27.9	1.73	1.13	119.7	10.4	115.1	10.7	0.50:0.30:0.21
20	151.5	48.6	1.46	1.21	144.4	9.2	144.3	11.5	0.50:0.33:0.17
20	155.9	12.8	1.50	0.32	120.7	25.2	117.2	37.0	0.44:0.43:0.14
20	163.6	16.9	1.57	1.00	0.9	8.8	5.7	9.2	0.47:0.33:0.20
20	178.0	1.2	1.23	0.37	78.9	20.3	77.0	14.7	0.51:0.31:0.17
20	216.4	1.4	1.11	0.86	21.6	8.1	28.4	11.3	0.49:0.30:0.21
21	133.5	41.2	1.40	0.72	16.9	12.1	18.5	10.6	0.49:0.33:0.18
21	170.6	16.8	1.07	0.90	93.0	10.0	94.1	15.4	0.47:0.37:0.17
21	191.8	11.0	1.06	4.43	40.9	2.8	41.0	2.9	0.66:0.24:0.10
21	209.1	3.2	1.56	0.77	55.9	14.8	12.3	25.9	0.49:0.30:0.21
21	212.9	12.6	1.55	1.72	162.7	3.9	162.8	3.9	0.63:0.22:0.15
21	231.2	25.2	1.51	4.57	177.7	3.9	178.8	2.4	0.57:0.29:0.14
22	136.8	49.5	1.19	1.03	119.4	8.2	126.4	13.2	0.46:0.38:0.17
22	182.0	55.5	1.70	1.88	126.1	4.6	126.3	4.7	0.56:0.23:0.20
22	217.8	− 0.8	1.31	0.42	71.0	33.3	69.2	27.0	0.42:0.33:0.25
23	155.9	53.5	1.48	1.79	115.6	4.8	115.8	4.8	0.60:0.22:0.19
23	166.4	37.1	1.31	0.97	134.5	10.2	126.3	15.9	0.44:0.37:0.19
23	171.3	14.0	1.20	1.57	117.5	6.4	101.2	15.7	0.48:0.37:0.15
23	180.5	6.0	1.29	1.26	81.2	13.5	72.8	11.3	0.57:0.25:0.18
23	209.5	34.3	1.65	1.99	152.5	2.8	152.3	2.8	0.68:0.17:0.15
23	214.3	31.8	1.48	0.27	18.6	32.6	87.7	32.5	0.47:0.35:0.18
24	119.1	18.6	1.28	1.06	157.3	9.6	162.8	11.1	0.46:0.32:0.23
24	139.6	2.5	1.20	0.61	48.9	6.7	49.0	6.7	0.58:0.29:0.13
24	171.1	17.6	1.52	0.83	78.7	11.2	77.0	10.3	0.50:0.28:0.22
24	179.6	65.0	1.08	0.71	35.4	3.7	35.2	3.9	0.55:0.23:0.22
24	205.0	12.0	1.36	0.61	129.2	16.1	129.2	16.6	0.46:0.33:0.20
24	217.1	33.8	1.11	0.99	12.8	15.0	34.9	21.8	0.54:0.28:0.18
24	217.1	57.5	1.70	0.75	9.8	14.3	11.6	10.3	0.52:0.28:0.20
25	142.5	31.3	1.33	1.70	42.4	6.1	41.1	7.2	0.50:0.34:0.15
25	149.5	43.7	1.16	0.81	42.3	7.8	38.9	9.9	0.45:0.35:0.20
25	186.0	3.6	1.19	0.61	23.6	12.3	47.3	15.3	0.48:0.32:0.20
25	196.7	36.5	1.48	0.64	103.1	7.7	102.9	8.6	0.47:0.33:0.20
25	231.9	43.6	1.20	0.30	91.9	15.3	75.1	16.9	0.44:0.36:0.21
26	160.3	53.5	1.18	0.33	110.6	23.2	111.5	25.1	0.38:0.34:0.28
26	171.7	24.2	1.10	0.78	48.5	8.3	47.3	8.1	0.46:0.29:0.24
26	206.0	6.5	1.43	0.75	38.1	8.7	46.7	12.0	0.46:0.36:0.18
26	224.2	61.6	1.57	1.52	106.8	3.1	106.8	4.3	0.51:0.33:0.16
26	245.6	40.3	1.55	0.58	4.2	8.0	3.5	4.7	0.49:0.26:0.25
27	184.1	52.7	1.18	0.37	124.4	28.7	125.7	28.7	0.40:0.35:0.25
27	204.0	13.7	1.16	0.98	19.6	5.8	19.9	8.2	0.50:0.36:0.13
27	221.8	52.4	1.55	1.12	52.3	7.2	54.5	6.6	0.54:0.26:0.20
27	225.5	57.3	1.52	0.54	63.3	14.4	45.6	14.6	0.45:0.37:0.18
27	231.2	15.4	1.56	1.01	60.8	14.0	53.2	18.7	0.43:0.35:0.22
28	138.6	14.8	1.54	1.22	135.6	8.1	139.3	7.1	0.63:0.21:0.15
28	177.0	43.1	1.54	1.10	45.7	6.2	45.2	6.1	0.52:0.25:0.23
28	192.3	10.7	1.36	0.51	155.8	15.8	137.9	12.9	0.49:0.32:0.19
28	212.0	61.7	1.15	1.33	131.7	5.5	135.9	7.6	0.61:0.23:0.16
30	138.1	7.9	1.56	0.85	39.7	8.4	41.0	9.5	0.51:0.32:0.17
30	217.2	10.5	1.67	0.99	147.5	6.5	146.9	6.8	0.56:0.23:0.21
30	234.0	21.8	1.60	0.61	91.3	12.6	107.0	16.9	0.46:0.33:0.21
31	176.5	38.3	1.28	0.37	132.6	17.9	134.4	16.6	0.46:0.34:0.19
31	202.7	21.3	1.08	0.74	64.1	6.0	65.3	5.3	0.56:0.24:0.20
32	177.4	32.7	1.18	2.11	46.6	4.5	46.1	4.6	0.59:0.23:0.18
33	126.6	20.3	1.44	1.15	109.6	5.6	112.8	5.7	0.52:0.29:0.20
33	157.8	20.7	1.59	1.07	15.6	13.0	12.7	11.9	0.41:0.33:0.26
33	183.6	11.0	1.08	0.69	111.3	4.1	111.2	3.7	0.54:0.25:0.21
34	162.3	5.3	1.28	0.60	108.6	11.9	153.9	25.5	0.44:0.37:0.19
34	228.5	8.1	1.22	0.71	178.7	8.2	156.5	17.2	0.45:0.34:0.21
34	234.5	10.7	1.24	0.88	55.9	7.4	53.5	8.3	0.48:0.29:0.24

Table C1 – *continued*

m	J2000 (°)				2D PA (°)		3D PA (°)		
	$\bar{\alpha}$	$\bar{\delta}$	$\bar{z}$	w_{71}	θ	γ_h	θ	γ_h	$a: b: c$
36	189.0	44.1	1.39	1.22	41.3	4.2	38.9	4.1	0.53:0.28:0.19
37	189.5	20.3	1.46	0.66	46.7	10.5	62.4	23.4	0.42:0.39:0.19
38	161.6	3.5	1.11	0.35	137.7	17.5	119.2	19.3	0.46:0.37:0.18
38	227.6	41.4	1.54	0.50	54.7	7.0	51.2	7.7	0.50:0.34:0.16
41	205.3	50.4	1.39	1.36	51.3	2.7	51.0	2.7	0.63:0.22:0.15
43	231.0	47.8	1.57	0.76	30.4	5.5	35.0	6.0	0.47:0.32:0.20
44	208.7	25.8	1.28	0.45	120.0	8.6	131.2	6.8	0.54:0.27:0.19
46	226.7	16.7	1.09	0.63	136.2	7.2	133.9	7.5	0.47:0.30:0.23
55	196.5	27.1	1.59	0.95	107.5	3.4	107.0	3.4	0.58:0.24:0.18
56	167.0	33.8	1.11	0.81	110.2	3.5	110.4	3.8	0.50:0.29:0.21
64	196.4	39.9	1.14	0.83	133.6	3.0	133.9	3.2	0.48:0.36:0.17
73	164.1	14.1	1.27	0.76	156.6	4.2	156.3	4.5	0.55:0.28:0.16

APPENDIX D: AXIAL DATA

Position angle data are axial [0°, 180°), more specifically 2-axial; 0° and 180° are equivalent. When analysing alignment, the orientation of the axis is important, but its direction (i.e. which end is the 'head' and which is the 'tail') is arbitrary and has no physical meaning. Axial data are synonymous with undirected data, unlike vectors which are directed.

Statistical analysis of directed data typically uses vector algebra (known as Fisher statistics). However, non-directed data cannot be treated as vectors. Fisher (1993) recommends a statistically valid solution for axial (2-axial) or, generally, *p*-axial data. First transform the angles to vector (circular) data as

$$\Theta \ [0°, \ 360°) = \begin{cases} 2 \times \theta & \text{for axial data} & [0°, \ 180°), \\ p \times \theta & \text{for } p\text{-axial data} & [0°, \ 360°/p), \end{cases} \quad \text{(D1)}$$

then analyse the data as required and back-transform the results. The final step, back-transformation, is generally required only to find direction (Fisher 1993). In the case of axial data, back-transformation is simply halving any resultant angles, e.g. to determine the direction of a mean resultant vector.

Hutsemékers et al. (2014) and Pelgrims (2016) test for alignment and, simultaneously, for orthogonality (which they describe as 'anti-alignment'), by converting from 2-axial θ [0°, 180°) to 4-axial θ_{4ax} [0°, 90°) data using

$$\theta_{4ax} \ [0°, \ 90°) = \text{mod}(\theta, 90°). \quad \text{(D2)}$$

For vector algebra, this is transformed to circular data Θ using equation (D1), specifically

$$\Theta_{4ax} \ [0°, \ 360°) = 4 \times \theta_{4ax}, \quad \text{(D3)}$$

where back-transformation, if required, would be to quarter any resultant angles.

Throughout this work we specify which construct of PA data we use, i.e. raw 2-axial (θ), circular 2-axial ($\Theta_{2ax} = 2\theta$, equation D1), 4-axial (θ_{4ax}, equation D2), or circular 4-axial ($\Theta_{4ax} = 4\theta_{4ax}$, equation D3).

APPENDIX E: PA UNCERTAINTIES

To estimate the measurement uncertainties in the LQG position angles we use bootstrap re-sampling with replacement (Efron 1979). For an LQG with *m* members, the data set consists of *m* observed quasar positions (right ascension, declination, and redshift). We construct *n* bootstrap LQGs, each with the same number of *m* members which are drawn at random from the original data set. Each draw is made from the entire data set, and each member is replaced in the data set before the next draw. Thus, each bootstrap LQG is likely to miss some members and have duplicates (or triplicates or more) of others.

These bootstraps are used to estimate the uncertainty on parameters (e.g. PAs) derived from the data set, without any assumption about the underlying population (Feigelson & Babu 2012). The position angle of each bootstrap LQG is calculated through the same 2D and 3D approaches used for the observed LQGs (Appendix C).

For a sample of *n* bootstrap LQGs we can determine the mean PA and its associated uncertainty. Using the linear mean $\bar{\theta} = (\sum_{i=1}^{n} \theta_i)/n$ is inappropriate for axial data [0°, 180°), where 0° and 180° are equivalent. For example, a sample of PAs centred on 0°, with around half in the range $0° \lesssim \theta \lesssim 10°$ and half in the range $170° \lesssim \theta \lesssim 180°$, has a linear mean $\bar{\theta} \sim 90°$ rather than the correct answer $\bar{\theta} \sim 0°$ (or, equivalently, $\bar{\theta} \sim 180°$).

Therefore, instead of linear mean, we calculate the 'circular' mean (Fisher 1993) of axial PAs. First, following Mardia & Jupp (2000), let

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} \cos 2\theta_i \ , \quad \bar{S} = \frac{1}{n} \sum_{i=1}^{n} \sin 2\theta_i, \quad \text{(E1)}$$

where θ_i is the PA of the *i*th bootstrap LQG and *n* is the total number of boostraps for this LQG. The factor of two accounts for the axial (rather than circular) nature of the data. The mean direction $\bar{\theta}$ is given by

$$\bar{\theta} = \begin{cases} \frac{1}{2} \arctan \left(\frac{\bar{S}}{\bar{C}} \right) & \text{if } \bar{C} \geq 0, \\ \frac{1}{2} \arctan \left(\frac{\bar{S}}{\bar{C}} \right) + \pi & \text{if } \bar{C} < 0, \end{cases} \quad \text{(E2)}$$

where the factor $\frac{1}{2}$ converts from vector algebra back to axial data (i.e. 'back-transformation').

For each LQG, to estimate the uncertainty in the mean $\bar{\theta}$, we calculate its confidence interval following Pelgrims (2016), who in turn follows Fisher (1993), and for each individual bootstrap *i* out of a total of *n* we define the residual as

$$\gamma_i = \frac{1}{2} \arctan \left(\frac{\sin(2(\theta_i - \bar{\theta}))}{\cos(2(\theta_i - \bar{\theta}))} \right), \quad \text{(E3)}$$

where θ_i is the PA of the *i*th bootstrap LQG. For $i = 1,..., n$ we sort the γ_i in ascending order to give an ordered list $\gamma_{(1)} \leq ... \leq \gamma_{(n)}$. To

(a) Example 1 - narrow LQG

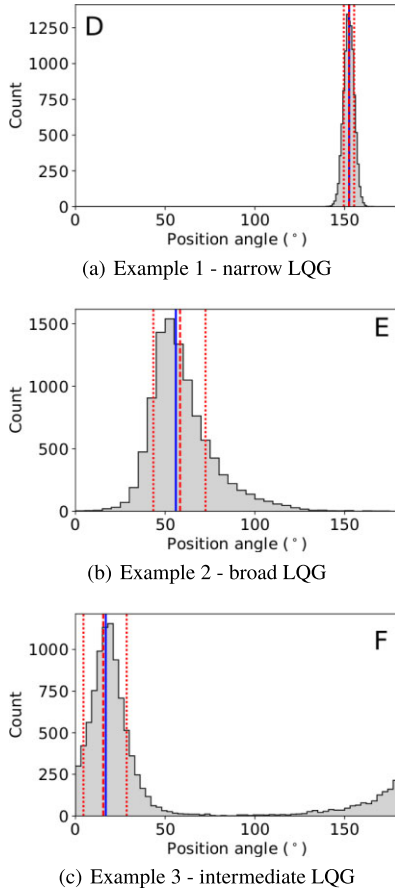(b) Example 2 - broad LQG

(c) Example 3 - intermediate LQG

Figure E1. LQG position angles from 10 000 bootstraps for three example LQGs, labelled D, E, F in Fig. B1, and visually classified as (a) narrow, (b) broad, and (c) intermediate morphology. All PA calculations use the 2D approach with no parallel transport. Dashed orange line shows the circular mean of the bootstraps $\bar{\theta}$, dotted orange lines show the 68 per cent confidence interval, solid blue line shows the observed PA.

determine the confidence interval at the $100(1 - \alpha)$ per cent level[8] we find the γ_i list elements at lower index l which is the integer part of $(n\alpha + 1)/2$ and upper index $u = n - l$. The confidence interval for $\bar{\theta}$ is then $[\bar{\theta} + \gamma_{(l+1)}, \bar{\theta} + \gamma_{(u)}]$. We calculate the confidence interval at the 68 per cent level ($\alpha = 0.32$) and define the half-width of the confidence interval (HWCI) as

$$\gamma_h = (\gamma_{(u)} - \gamma_{(l+1)})/2. \qquad (E4)$$

For each of our sample of 71 LQGs, we create $n = 10\,000$ bootstraps, calculate their PAs θ_i using the 2D or 3D approach (Appendix C), and then determine the mean $\bar{\theta}$ and confidence interval. Fig. E1 shows the distribution of bootstrap PAs θ_i calculated using the 2D approach for three example LQGs. The circular mean of the bootstraps $\bar{\theta}$ (dashed orange line) generally agrees well with the observed PA (solid blue line). As expected, the 68 per cent confidence interval (Fig. E1, dotted orange lines) is smaller for (a) a 'narrow' LQG than (b) a 'broad' sheet-like LQG. Example (c) illustrates why the linear mean is inappropriate, since the distribution of the bootstrap PAs may 'wrap' from 180° back to 0°.

The HWCIs for 10 000 bootstraps of our sample of 71 LQGs, with PAs θ_i calculated using both the 2D and 3D approaches, are shown as

error bars in Fig. C3. The mean (median) HWCI for the 2D approach is ∼10° (∼8°), and for the 3D approach it is ∼11° (∼9°).

APPENDIX F: APPLYING THE S TEST

F1 Nearest neighbours free parameter

The S test quantifies the coherence of PA alignment by measuring the dispersion of groups of n_v nearest neighbours, where n_v is a free parameter. We explore a range of n_v values; we do not choose a specific value. For each LQG, its nearest neighbours can be determined either in two dimensions (angular separation) or three dimensions (proper separation).

In 2D, nearest neighbours are identified by calculating the angular separation θ on the celestial sphere between LQG 1 and LQG 2 as

$$\theta = \cos^{-1}[\sin\delta_1 \sin\delta_2 + \cos\delta_1 \cos\delta_2 \cos(\alpha_1 - \alpha_2)], \qquad (F1)$$

where α_1 and δ_1 (α_2 and δ_2) are the right ascension and declination of the centroids of LQG 1 (2), respectively. For each LQG, its nearest neighbours are those n_v LQGs separated from it by the smallest angular distances.

In 3D, nearest neighbours are identified by calculating the 3D proper positions (x, y, z) of each LQG centroid. For each LQG, its nearest neighbours are those n_v LQGs separated from it by the smallest proper distances.

The 2D and 3D approaches of identifying nearest neighbours will return different groups of LQGs. When these groups are used to compute the S statistic S_D we find that the values calculated using the two approaches are remarkably consistent. This is probably due to the geometry of the 3D survey volume; our redshift restriction of $1.0 \leq z \leq 1.8$ yields a 'shell' of LQGs of finite thickness. At low n_v the 3D approach may find neighbours in the radial direction, but at high n_v it can only find them tangentially (on the sky), like the 2D approach. Since the two approaches give very similar results, and the 3D approach is more physically motivated, we use the 3D approach of identifying nearest neighbours in this work.

F2 Estimating significance level

The S test yields an S_D statistic for each number of nearest neighbours n_v assessed. The significance level of these values compared to randomness cannot be evaluated analytically, due to: overlaps between groups of nearest neighbours, deviation of the S_D distribution from normality (particularly for small n and when $n_v \sim n$), and the dependence of parallel transport corrections on the precise location of the objects involved. Therefore, numerical simulations are required.

Hutsemékers (1998) and Hutsemékers & Lamy (2001) generate random samples by shuffling the observed PAs randomly between objects, while keeping their positions fixed. This has the effect of erasing any correlation between PAs and positions. But, as Jain et al. (2004) note, the shuffling method is unable to test for global alignment.

We have no reason to expect LQG PAs to be correlated with LQG positions; alignment could be global. Therefore, we generate random samples from a uniform distribution, keeping the 3D LQG positions fixed. We generate $n = 71$ PAs, randomly drawn from a uniform distribution. These are generated in the ranges [0°, 180°) for 2-axial PAs and [0°, 90°) for 4-axial PAs. Note that initially generating 4-axial PAs in the range [0°, 180°), then converting to [0°, 90°) after parallel transport and before evaluating the S test is equivalent.

The significance level (SL) of the S test is defined as the percentage of simulations that have an S_D statistic at least as extreme as the one

---

[8]Confidence level $= 1 - \alpha$, where α is the significance level.

from our observations (Pelgrims 2016). It is computed by comparing the statistic of our observations ($S_{D,\,obs}$) with the statistic from a large number of numerical simulations ($S_{D,\,sim}$).

For 2-axial PAs we find that $S_{D,\,obs}$ is in the low $S_{D,\,sim}$ (left-hand) tail of the distribution. To interpret this, consider the potential bimodality of the LQG PA distribution, with the peaks separated by ∼90°. This orthogonality leads to a large dispersion, and yields a small value of S_D. Therefore, it is legitimate to interpret a result in the left-hand tail as a potentially orthogonal signal, with the SL being the probability that $S_{D,\,sim} < S_{D,\,obs}$. This can be tested by also evaluating the S test for 4-axial PAs.

Note that two modes of a possible combined 'alignment plus orthogonality' signal would tend to erase any signal, reducing the power of this test. However, a residual signal in the left (right) tail indicates the orthogonality (alignment) mode dominates.

When we convert 2-axial PAs to 4-axial, we simultaneously test for alignment and orthogonality by combining the modes, so an 'alignment plus orthogonality' signal will manifest as alignment only. The SL is therefore the proportion of numerical simulations with statistic S_D higher than that observed. Indeed, for 4-axial PAs we find that $S_{D,\,obs}$ is in the high $S_{D,\,sim}$ (right-hand) tail of the distribution, with the SL being the probability that $S_{D,\,sim} > S_{D,\,obs}$. Note that an 'alignment only' signal would also be in the right-hand tail, but would be differentiated by its 2-axial result.

The S_D distributions of 2-axial and 4-axial numerical simulations differ, particularly as n_v approaches n. Therefore, we generate separate numerical simulations for each, and calculate the significance levels as

$$\mathrm{SL}_{2\mathrm{ax}} = P(S_{D,sim(2ax)} < S_{D,obs(2ax)}), \qquad (\mathrm{F2})$$

$$\mathrm{SL}_{4\mathrm{ax}} = P(S_{D,sim(4ax)} > S_{D,obs(4ax)}), \qquad (\mathrm{F3})$$

where P indicates probability, *2ax* and *4ax* indicate 2-axial and 4-axial PAs, and *sim* and *obs* indicate simulations and observations, respectively.

APPENDIX G: LQGS IN 3D PROPER SPACE

The 71 LQG positions in 3D proper space, and their orientation as determined by the 3D method, are illustrated in Fig. G1, with additional perspectives shown in Fig. G2.
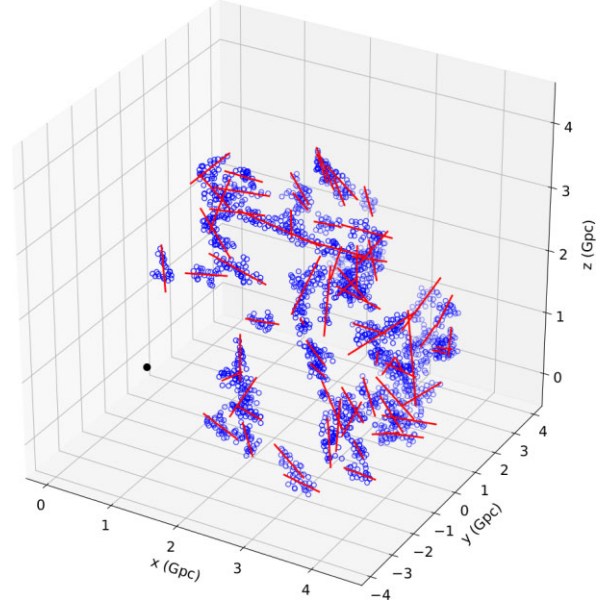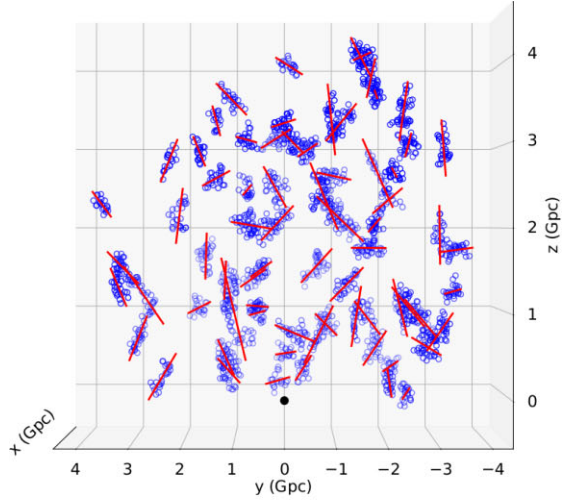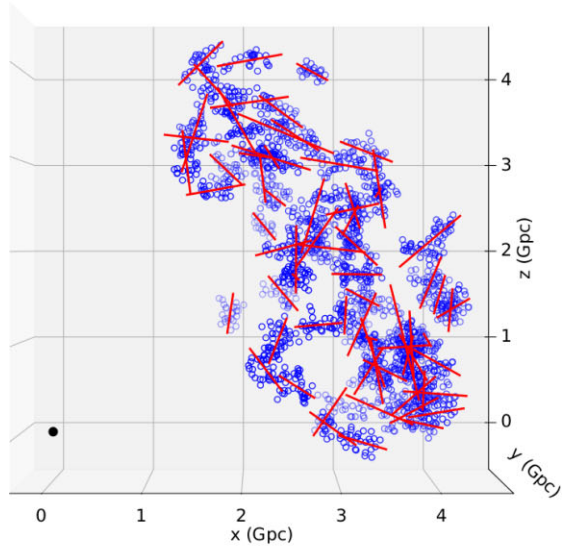

Figure G1. LQGs in 3D proper space, showing LQG quasars (blue circles) and LQG major axes (red lines). Quasar markers are shaded to give the appearance of depth, with lighter shades representing more distant quasars. Our location at (x, y, z) coordinates (0, 0, 0) is indicated by a black dot. See Fig. G2 for three alternative, orthogonal perspectives.

(a) *x*-axis orthogonal to the page

(b) *y*-axis orthogonal to the page

(c) *z*-axis orthogonal to the page

Figure G2. As Fig. G1 but viewed with the (a) *x*-axis, (b) *y*-axis, and (c) *z*-axis orthogonal to the page.
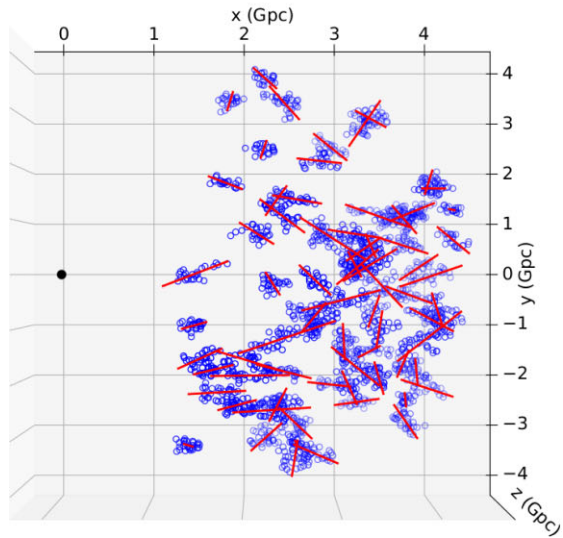
APPENDIX H: PARALLEL TRANSPORT DESTINATION

Fig. H1 shows LQG PAs after parallel transport to the location of each of the 71 LQGs. Also shown, double Gaussian fit (blue).
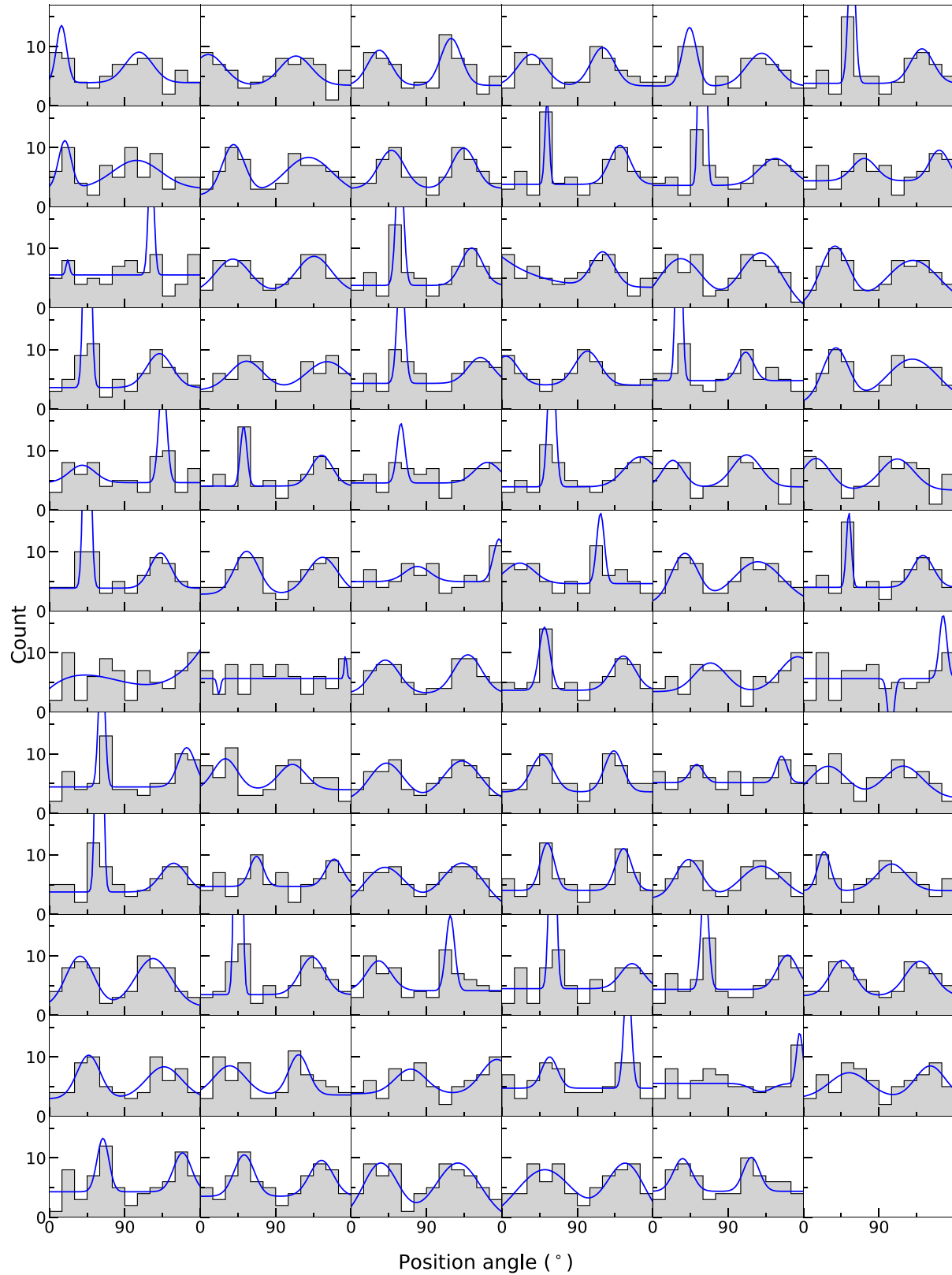
Figure H1. LQG position angles after parallel transport to the location of each LQG; 15° bins. Solid blue lines are double Gaussian fit. The bimodal distribution is generally robust to parallel transport destination.

This paper has been typeset from a TEX/LATEX file prepared by the author.